

ISSN 2156-5570(Online)

ISSN 2158-107X(Print)

Editorial Preface

From the Desk of Managing Editor...

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 14 Issue 8 August 2023
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University

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

Associate Editors

Alaa Sheta

Southern Connecticut State University

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

Domenico Ciuonzo

University of Naples, Federico II, Italy

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

Dorota Kaminska

Lodz University of Technology

Domain of Research: Artificial Intelligence, Virtual Reality

Elena Scutelnicu

"Dunarea de Jos" University of Galati

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

In Soo Lee

Kyungpook National University

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

Krassen Stefanov

Professor at Sofia University St. Kliment Ohridski

Domain of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, e-Learning Tools, Educational Systems Design

Renato De Leone

Università di Camerino

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

Xiao-Zhi Gao

University of Eastern Finland

Domain of Research: Artificial Intelligence, Genetic Algorithms

CONTENTS

Paper 1: An Empirical Internet Protocol Network Intrusion Detection using Isolation Forest and One-Class Support Vector Machines

Authors: Gerard Shu Fuhnwi, Victoria Adedoyin, Janet O. Agbaje

PAGE 1 – 6

Paper 2: Ensemble Security and Multi-Cloud Load Balancing for Data in Edge-based Computing Applications

Authors: Raghunadha Reddi Dornala

PAGE 7 – 13

Paper 3: Converting Data for Spiking Neural Network Training

Authors: Erik Sadvovsky, Maros Jakubec, Roman Jarina

PAGE 14 – 20

Paper 4: A Secure and Scalable Behavioral Dynamics Authentication Model

Authors: Idowu Dauda Oladipo, Joseph Bamidele Awotunde, Mathew Nicho, Jemima Omotola Buari, Muyideen Abdulaheem, Tarek Gaber

PAGE 21 – 32

Paper 5: Visualization of AI Systems in Virtual Reality: A Comprehensive Review

Authors: Medet Inkarbekov, Rosemary Monahan, Barak A. Pearlmutter

PAGE 33 – 42

Paper 6: Symbol Detection in a Multi-class Dataset Based on Single Line Diagrams using Deep Learning Models

Authors: Hina Bhanbhro, Yew Kwang Hooi, Worapan Kusakunniran, Zaira Hassan Amur

PAGE 43 – 56

Paper 7: The Spatial Distribution of Atmospheric Water Vapor Based on Analytic Hierarchy Process and Genetic Algorithm

Authors: Fengjun Wei, Chunhua Liu, Rendong Guo, Xin Li, Jilei Hu, Chuanxun Che

PAGE 57 – 68

Paper 8: Detection of Tuberculosis Based on Hybridized Pre-Processing Deep Learning Method

Authors: Mohamed Ahmed Elashmawy, Irraivan Elamvazuthi, Lila Iznita Izhar, Sivajothi Paramasivam, Steven Su

PAGE 69 – 76

Paper 9: Automated CAD System for Early Stroke Diagnosis: Review

Authors: Izzatul Husna Azman, Norhashimah Mohd Saad, Abdul Rahim Abdullah, Rostam Affendi Hamzah, Adam Samsudin, Shaarmila A/P Kandaya

PAGE 77 – 83

Paper 10: The Current State of Blockchain Consensus Mechanism: Issues and Future Works

Authors: Shadab Alam

PAGE 84 – 94

Paper 11: A Novel Approach for Identification of Figurative Language Types in Devanagari Scripted Languages

Authors: Jatinderkumar R. Saini, Preeti Sagar, Hema Gaikwad

PAGE 95 – 103

Paper 12: Machine Learning Model for Automated Assessment of Short Subjective Answers

Authors: Zaira Hassan Amur, Yew Kwang Hooi, Hina Bhanbro, Mairaj Nabi Bhatti, Gul Muhammad Soomro

PAGE 104 – 112

Paper 13: Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach

Authors: Helmi Imaduddin, Fiddin Yusufida A'la, Yusuf Sulisty Nugroho

PAGE 113 – 117

Paper 14: The Medical Image Denoising Method Based on the CycleGAN and the Complex Shearlet Transform

Authors: ChunXiang Liu, Jin Huang, Muhammad Tahir, Lei Wang, Yuwei Wang, Faiz Ullah

PAGE 118 – 125

Paper 15: Comparing Scrum Maturity of Digital and Business Process Reengineering Groups: A Case Study at an Indonesia's State-Owned Bank

Authors: Gloria Saripah Patara, Teguh Raharjo

PAGE 126 – 134

Paper 16: Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities

Authors: Bridget Ogheneovo Malasowe, Maureen Ifeanyi Akazue, Ejaita Abugor Okpako, Fidelis Obukohwo Aghware, Arnold Adimabua Ojugo, Deborah Voke Ojie

PAGE 135 – 142

Paper 17: A Yolo-based Violence Detection Method in IoT Surveillance Systems

Authors: Hui Gao

PAGE 143 – 149

Paper 18: Towards Automated Evaluation of the Quality of Educational Services in HEIs

Authors: Silvia Gaffandzhieva, Rositsa Doneva, Mariya Zhekova, George Pashev

PAGE 150 – 165

Paper 19: Machine-Learning-based User Behavior Classification for Improving Security Awareness Provision

Authors: Alaa Al-Mashhour, Areej Alhogail

PAGE 166 – 178

Paper 20: Collateral Circulation Classification Based on Cone Beam Computed Tomography Images using ResNet18 Convolutional Neural Network

Authors: Nur Hasanah Ali, Abdul Rahim Abdullah, Norhashimah Mohd Saad, Ahmad Sobri Muda

PAGE 179 – 185

Paper 21: An Enhanced Algorithm of Improved Response Time of ITS-G5 Protocol

Authors: Kawtar Jellid, Tomader Mazri

PAGE 186 – 194

Paper 22: Design and Application of an Automatic Scoring System for English Composition Based on Artificial Intelligence Technology

Authors: Fengqin Zhang

PAGE 195 – 205

Paper 23: An Efficient Deep Learning with Optimization Algorithm for Emotion Recognition in Social Networks

Authors: Ambika G N, Yeresime Suresh

PAGE 206 – 215

Paper 24: Improved Drosophila Visual Neural Network Application in Vehicle Target Tracking and Collision Warning

Authors: Jianyi Wu

PAGE 216 – 226

Paper 25: Earth Observation Satellite: Big Data Retrieval Method with Fuzzy Expression of Geophysical Parameters and Spatial Features

Authors: Kohei Arai

PAGE 227 – 234

Paper 26: Virtual Route Guide Chatbot Based on Random Forest Classifier

Authors: Puspa Miladin Nuraida Safitri A. Basid, Fajar Rohman Hariri, Fresy Nugroho, Ajib Hanani, Firman Jati Pamungkas

PAGE 235 – 241

Paper 27: Approaches and Tools for Quality Assurance in Distance Learning: State-of-play

Authors: Silvia Gaffandzhieva, Rositsa Doneva, Senthil Kumar Jagatheesaperumal

PAGE 242 – 253

Paper 28: Efficient Parameter Estimation in Image Processing using a Multi-Agent Hysteretic Q-Learning Approach

Authors: Issam QAFFOU

PAGE 254 – 262

Paper 29: Design and Implementation of an IoT Control and Monitoring System for the Optimization of Shrimp Pools using LoRa Technology

Authors: José M. Pereira Pontón, Verónica Ojeda, Víctor Asanza, Leandro L. Lorente-Leyva, Diego H. Peluffo-Ordóñez

PAGE 263 – 272

Paper 30: An Overview of Vision Transformers for Image Processing: A Survey

Authors: Ch.Sita Kameswari, Kavitha J, T. Srinivas Reddy, Balaswamy Chinthaguntla, Senthil Kumar Jagatheesaperumal, Silvia Gaffandzhieva, Rositsa Doneva

PAGE 273 – 289

Paper 31: Multimodal Contactless Architecture for Upper Limb Virtual Rehabilitation

Authors: Emilio Valdivia-Cisneros, Elizabeth Vidal, Eveling Castro-Gutierrez

PAGE 290 – 295

Paper 32: Attitude Synchronization and Stabilization for Multi-Satellite Formation Flying with Advanced Angular Velocity Observers

Authors: Belkacem Kada, Khalid Munawar, Muhammad Shafique Shaikh

PAGE 296 – 303

Paper 33: Hussein Search Algorithm: A Novel Efficient Searching Algorithm in Constant Time Complexity

Authors: Omer H Abu El Hajjia, Arwa H. F. Zabian

PAGE 304 – 309

Paper 34: Testing the Usability of Serious Game for Low Vision Children

Authors: Nurul Izzah Othman, Hazura Mohamed, Nor Azan Mat Zin

PAGE 310 – 317

Paper 35: Cybersecurity Advances in SCADA Systems

Authors: Bakil Al-Muntaser, Mohamad Afendee Mohamed, Ammar Yaseen Tuama, Imran Ahmad Rana

PAGE 318 – 328

Paper 36: Model Classification of Fire Weather Index using the SVM-FF Method on Forest Fire in North Sumatra, Indonesia

Authors: Darwis Robinson Manalu, Opim Salim Sitompul, Herman Mawengkang, Muhammad Zarlis

PAGE 329 – 337

Paper 37: Prediction of Cryptocurrency Price using Time Series Data and Deep Learning Algorithms

Authors: Michael Nair, Mohamed I. Marie, Laila A. Abd-Elmegid

PAGE 338 – 347

Paper 38: Advances in Value-based, Policy-based, and Deep Learning-based Reinforcement Learning

Authors: Haewon Byeon

PAGE 348 – 354

Paper 39: Scalable Blockchain Architecture: Leveraging Hybrid Shard Generation and Data Partitioning

Authors: Praveen M Dhulavvagol, Prasad M R, Niranjan C Kundur, Jagadisha N, S G Totad

PAGE 355 – 363

Paper 40: Detection of Herd Pigs Based on Improved YOLOv5s Model

Authors: Jianquan LI, Xiao WU, Yuanlin NING, Ying YANG, Gang LIU, Yang MI

PAGE 364 – 370

Paper 41: The Impact of Cyber Security on Preventing and Mitigating Electronic Crimes in the Jordanian Banking Sector

Authors: Tamer Bani Amer, Mohammad Ibrahim Ahmed Al-Omar

PAGE 371 – 380

Paper 42: Research on the Local Path Planning for Mobile Robots based on PRO-Dueling Deep Q-Network (DQN) Algorithm

Authors: Yaoyu Zhang, Caihong Li, Guosheng Zhang, Ruihong Zhou, Zhenying Liang

PAGE 381 – 387

Paper 43: Prostate Cancer Detection and Analysis using Advanced Machine Learning

Authors: Mowafaq Salem Alzboon, Mohammad Subhi Al-Batah

PAGE 388 – 396

Paper 44: Application of Improved Ant Colony Algorithm Integrating Adaptive Parameter Configuration in Robot Mobile Path Design

Authors: Jinli Han

PAGE 397 – 407

Paper 45: Simulation of Logistics Frequent Path Data Mining Based on Statistical Density

Authors: Fengju Hou

PAGE 408 – 413

Paper 46: Simulation Analysis of Hydraulic Control System of Engineering Robot Arm Based on ADAMS

Authors: Haiqing Wu

PAGE 414 – 420

Paper 47: Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging

Authors: Muneer Majid, Yonis Gulzar, Shahnawaz Ayoub, Farhana Khan, Faheem Ahmad Reegu, Mohammad Shuaib Mir, Wassim Jaziri, Arjumand Bano Soomro

PAGE 421 – 432

Paper 48: A Hybrid Metaheuristic Model for Efficient Analytical Business Prediction

Authors: Marischa Elveny, Mahyuddin K. M Nasution, Rahmad B. Y Syah

PAGE 433 – 440

Paper 49: A Mechanism for Bitcoin Price Forecasting using Deep Learning

Authors: Karamath Ateeq, Ahmed Abdelrahim Al Zarooni, Abdur Rehman, Muhammd Adna Khan

PAGE 441 – 448

Paper 50: Research on Improving Piano Performance Evaluation Method in Piano Assisted Online Education

Authors: Huayi Qi, Chunhua She

PAGE 449 – 459

Paper 51: Methodological Insights Towards Leveraging Performance in Video Object Tracking and Detection

Authors: Divyaprabha, M. Z Kurian

PAGE 460 – 474

Paper 52: Pairwise Test Case Generation using (1+1) Evolutionary Algorithm for Software Product Line Testing

Authors: Sharafeldin Kabashi Khatir, Rabatul Aduni Binti Sulaiman, Mohammed Adam Kunna Azrag, Jasni Mohamad Zain, Julius Beneoluchi Odili, Samer Ali Al-Shami

PAGE 475 – 483

Paper 53: Campus Network Intrusion Detection Based on Gated Recurrent Neural Network and Domain Generation Algorithm

Authors: Qi Rong, Guang Zhao

PAGE 484 – 492

Paper 54: Dynamic Modelling of Hand Grasping and Wrist Exoskeleton: An EMG-based Approach

Authors: Mohd Safirin Bin Karis, Hyreil Anuar Bin Kasdirin, Norafizah Binti Abas, Muhammad Noorazlan Shah Bin Zainudin, Sufri Bin Muhammad, Mior Muhammad Nazmi Firdaus Bin Mior Fadzil

PAGE 493 – 499

Paper 55: Research on Semantic Segmentation Method of Remote Sensing Image Based on Self-supervised Learning

Authors: Wenbo Zhang, Achuan Wang

PAGE 500 – 508

Paper 56: Mechatronics Design and Robotic Simulation of Serial Manipulators to Perform Automation Tasks in the Avocado Industry

Authors: Carlos Paredes, Ricardo Palomares, Josmell Alva, José Cornejo

PAGE 509 – 517

Paper 57: Integrating Transfer Learning and Deep Neural Networks for Accurate Medical Disease Diagnosis from Multi-Modal Data

Authors: Chamandeep Kaur, Abdul Rahman Mohammed Al-Ansari, Taviti Naidu Gongada, K. Aanandha Saravanan, Divvela Srinivasa Rao, Ricardo Fernando Cosio Borda, R. Manikandan

PAGE 518 – 528

Paper 58: An Integrated Instrument for Measuring Science, Technology, Engineering, and Mathematics: Digital Educational Game Acceptance and Player Experience

Authors: Husna Hafiza R. Azami, Roslina Ibrahim, Suraya Masrom, Rasimah Che Mohd Yusoff, Suraya Yaacob

PAGE 529 – 539

Paper 59: Human-object Behavior Analysis Based on Interaction Feature Generation Algorithm

Authors: Qing Ye, Xiuju Xu, Rui Li

PAGE 540 – 549

Paper 60: A Proposed Framework for Context-Aware Semantic Service Provisioning

Authors: Wael Haider, Hatem Abdelkader, Amira Abdelwahab

PAGE 550 – 559

Paper 61: Impact of the Use of the Video Game SimCity on the Development of Critical Thinking in Students: A Quantitative Experimental Approach

Authors: Jorge Luis Torres-Loayza, Grunilda Telma Reymer-Morales, Benjamín Maraza-Quispe

PAGE 560 – 567

Paper 62: An Automated Medical Image Segmentation Framework using Deep Learning and Variational Autoencoders with Conditional Neural Networks

Authors: Dustakar Surendra Rao, L. Koteswara Rao, Bhagyaraju Vipparthi

PAGE 568 – 578

Paper 63: Estimating Probability Values Based on Naïve Bayes for Fuzzy Random Regression Model

Authors: Hamijah Mohd Rahman, Nureize Arbaiy, Chuah Chai Wen, Pei-Chun Lin

PAGE 579 – 584

Paper 64: A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data

Authors: Hairani Hairani, Dadang Priyanto

PAGE 585 – 590

Paper 65: An Ensemble Load Balancing Algorithm to Process the Multiple Transactions Over Banking

Authors: Raghunadha Reddi Dornala

PAGE 591 – 596

Paper 66: Genetic Approach for Improved Prediction of Adaptive Learning Activities in Intelligent Tutoring System

Authors: Fatima-Zohra Hibbi, Otman Abdoun, El Khatir Haimoudi

PAGE 597 – 603

Paper 67: Algorithm for Skeleton Action Recognition by Integrating Attention Mechanism and Convolutional Neural Networks

Authors: Jianhua Liu

PAGE 604 – 613

Paper 68: A Population-based Plagiarism Detection using DistilBERT-Generated Word Embedding

Authors: Yuqin JING, Ying LIU

PAGE 614 – 624

Paper 69: Enhancing Startup Efficiency: Multivariate DEA for Performance Recognition and Resource Optimization in a Dynamic Business Landscape

Authors: K. N. Preethi, Yousef A. Baker El-Ebiary, Esther Rosa Saenz Arenas, Kathari Santosh, Ricardo Fernando Cosio Borda, Jorge L. Javier Vidalón, Anuradha. S, R. Manikandan

PAGE 625 – 635

Paper 70: Design and Improvement of New Industrial Robot Mechanism Based on Innovative BP-ARIMA Combined Model

Authors: Yuanyuan Liu

PAGE 636 – 642

Paper 71: A Proposed Approach for Monkeypox Classification

Authors: Luong Hoang Huong, Nguyen Hoang Khang, Le Nhat Quynh, Le Huu Thang, Dang Minh Canh, Ha Phuoc Sang

PAGE 643 – 651

Paper 72: CryptoScholarChain: Revolutionizing Scholarship Management Framework with Blockchain Technology

Authors: Jadhav Swati, Pise Nitin

PAGE 652 – 659

Paper 73: The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business

Authors: Mohd Shahrizan Abd Rahman, Nor Azliana Akmal Jamaludin, Zuraini Zainol, Tengku Mohd Tengku Sembok

PAGE 660 – 667

Paper 74: Deep Learning Enhanced Internet of Medical Things to Analyze Brain Computed Tomography Images of Stroke Patients

Authors: Batyrkhan Omarov, Azhar Tursynova, Meruert Uzak

PAGE 668 – 676

Paper 75: Chatbot Program for Proposed Requirements in Korean Problem Specification Document

Authors: Young Yun Baek, Soojin Park, Young B. Park

PAGE 677 – 681

Paper 76: Applying Artificial Intelligence and Computer Vision for Augmented Reality Game Development in Sports

Authors: Nurlan Omarov, Bakhytzhan Omarov, Axaule Baibaktina, Bayan Abilmazhinova, Tolep Abdimukhan, Bauyrzhan Doskarayev, Akzhan Adilzhan

PAGE 682 – 689

Paper 77: PMG-Net: Electronic Music Genre Classification using Deep Neural Networks

Authors: Yuemei Tang

PAGE 690 – 698

Paper 78: Automatic Layout Algorithm for Graphic Language in Visual Communication Design

Authors: Xiaofang Liao, Xinqian Hu

PAGE 699 – 708

Paper 79: Smart Sensor Signal-Assisted Behavioral Model and Control of Live Interaction in Digital Media Art

Authors: Pujie Li, Shi Bai

PAGE 709 – 718

Paper 80: Research on Strategic Decision Model of Human Resource Management based on Biological Neural Network

Authors: Ke Xu

PAGE 719 – 729

Paper 81: Multimodal Deep Learning Approach for Real-Time Sentiment Analysis in Video Streaming

Authors: Tejashwini S. G, Aradhana D

PAGE 730 – 736

Paper 82: 3D Magnetic Resonance Image Denoising using Wasserstein Generative Adversarial Network with Residual Encoder-Decoders and Variant Loss Functions

Authors: Hanaa A. Sayed, Anoud A. Mahmoud, Sara S. Mohamed

PAGE 737 – 746

Paper 83: A Framework for Patient-Centric Medical Image Management using Blockchain Technology

Authors: Abdulaziz Aljaloud

PAGE 747– 757

Paper 84: An Ensemble Learning Approach for Multi-Modal Medical Image Fusion using Deep Convolutional Neural Networks

Authors: Andino Maselena, D. Kavitha, Koudegai Ashok, Mohammed Saleh Al Ansari, Nimmati Satheesh, R. Vijaya Kumar Reddy

PAGE 758 – 769

Paper 85: Segmentation of Breast Cancer on Ultrasound Images using Attention U-Net Model

Authors: Sara LAGHMATI, Khadija HICHAM, Bouchaib CHERRADI, Soufiane HAMIDA, Amal TMIRI

PAGE 770 – 778

Paper 86: New Real Dataset Creation to Develop an Intelligent System for Predicting Chemotherapy Protocols

Authors: Houda AIT BRAHIM, Mariam BENLLARCH, Nada BENHIMA, Salah EL-HADAJ, Abdelmoutalib METRANE, Ghizlane BELBARAKA

PAGE 779 – 785

Paper 87: Presenting a Novel Method for Identifying Communities in Social Networks Based on the Clustering Coefficient

Authors: Zhihong HE, Tao LIU

PAGE 786 – 794

Paper 88: Motor Imagery EEG Signals Marginal Time Coherence Analysis for Brain-Computer Interface

Authors: Md. Sujan Ali, Jannatul Ferdous

PAGE 795 – 805

Paper 89: Systematic Review for Phonocardiography Classification Based on Machine Learning

Authors: Abdullah Altaf, Hairulnizam Mahdin, Awais Mahmood Alive, Mohd Izuan Hafez Ninggal, Abdulrehman Altaf, Irfan Javid

PAGE 806 – 817

Paper 90: A Hybrid Classification Approach of Network Attacks using Supervised and Unsupervised Learning

Authors: Rahaf Hamoud R. Al-Ruwaili, Osama M. Ouda

PAGE 818 – 828

Paper 91: Violent Physical Behavior Detection using 3D Spatio-Temporal Convolutional Neural Networks

Authors: Xiuhong Xu, Zhongming Liao, Zhaosheng Xu

PAGE 829 – 836

Paper 92: Construction of VR Video Quality Evaluation Model Based on 3D-CNN

Authors: Hongxia Zhao, Li Huang

PAGE 837 – 845

Paper 93: Design Strategy and Application of Headwear with National Characteristics Based on Information Visualization Technology

Authors: Ting Zhang

PAGE 846 – 856

Paper 94: SLAM Mapping Method of Laser Radar for Tobacco Production Line Inspection Robot Based on Improved RBPF

Authors: Zhiyuan Liang, Pengtao He, Wenbin Liang, Xiaolei Zhao, Bin Wei

PAGE 857 – 866

Paper 95: Visual Image Feature Recognition Method for Mobile Robots Based on Machine Vision

Authors: Minghe Hu, Jiancang He

PAGE 867 – 874

Paper 96: Explore Chinese Energy Commodity Prices in Financial Markets using Machine Learning

Authors: Yu Cui, Tianhao Ma

PAGE 875 – 880

Paper 97: Research on the Application of Multi-Objective Algorithm Based on Tag Eigenvalues in e-Commerce Supply Chain Forecasting

Authors: Man Huang, Jie Lian

PAGE 881 – 891

Paper 98: Construction and Application of Automatic Scoring Index System for College English Multimedia Teaching Based on Neural Network

Authors: Hui Dong, Ping Wei

PAGE 892 – 900

Paper 99: Design of a Decentralized AI IoT System Based on Back Propagation Neural Network Model

Authors: Xiaomei Zhang

PAGE 901 – 909

Paper 100: Black Widow Optimization Algorithm for Virtual Machines Migration in the Cloud Environments

Authors: Chuang Zhou

PAGE 910 – 916

Paper 101: Towards Secure Blockchain-enabled Cloud Computing: A Taxonomy of Security Issues and Recent Advances

Authors: Shengli LIU

PAGE 917 – 926

Paper 102: Research on Enterprise Supply Chain Anti-Disturbance Management Based on Improved Particle Swarm Optimization Algorithm

Authors: Tongqing Dai

PAGE 927 – 936

Paper 103: Automated Analysis of Job Market Demands using Large Language Model

Authors: Myo Thida

PAGE 937 – 946

Paper 104: Decentralized Management of Medical Test Results Utilizing Blockchain, Smart Contracts, and NFTs

Authors: Quy T. L, Khanh H. V, Huong H. L, Khiem H. G, Phuc T. N, Ngan N. T. K, Triet M. N, Bang L. K, Trong D. P. N, Hieu M. D, Bao Q. T, Khoa D. T

PAGE 947 – 957

Paper 105: Leveraging Blockchain, Smart Contracts, and NFTs for Streamlining Medical Waste Management: An Examination of the Vietnamese Healthcare Sector

Authors: Triet M. N, Khanh H. V, Huong H. L, Khiem H. G, Phuc T. N, Ngan N. T. K, Quy T. L, Bang L. K, Trong D. P. N, Hieu M. D, Bao Q. T, Khoa D. T, Anh T. N

PAGE 958 – 970

Paper 106: A Novel Dual Confusion and Diffusion Approach for Grey Image Encryption using Multiple Chaotic Maps

Authors: S Phani Praveen, V Sathiya Suntharam, S Ravi, U. Harita, Venkata Nagaraju Thatha, D Swapna

PAGE 971 – 984

Paper 107: Implementing a Blockchain, Smart Contract, and NFT Framework for Waste Management Systems in Emerging Economies: An Investigation in Vietnam

Authors: Khiem H. G, Khanh H. V, Huong H. L, Quy T. L, Phuc T. N, Ngan N. T. K, Triet M. N, Bang L. K, Trong D. P. N, Hieu M. D, Bao Q. T, Khoa D. T

PAGE 985 – 996

Paper 108: Deep Learning-based Sentence Embeddings using BERT for Textual Entailment

Authors: Mohammed Alsuhaibani

PAGE 997 – 1004

Paper 109: An Approach of Test Case Generation with Software Requirement Ontology

Authors: Adisak Intana, Kuljaree Tantayakul, Kanjana Laosen, Suraiya Charoenreh

PAGE 1005 – 1014

Paper 110: Eligible Personal Loan Applicant Selection using Federated Machine Learning Algorithm

Authors: Mehrin Anannya, Most. Shahera Khatun, Md. Biplob Hosen, Sabbir Ahmed, Md. Farhad Hossain, M. Shamim Kaiser

PAGE 1015 – 1024

Paper 111: A Low-Cost Wireless Sensor System for Power Quality Management in Single-Phase Domestic Networks

Authors: Cristian A. Aldana B, Edison F. Montenegro A

PAGE 1025 – 1036

Paper 112: A Novel Convolutional Neural Network Architecture for Pollen-Bearing Honeybee Recognition

Authors: Thi-Nhung Le, Thi-Minh-Thuy Le, Thi-Thu-Hong Phan, Huu-Du Nguyen, Thi-Lan Le

PAGE 1037 – 1044

Paper 113: Tomato Disease Recognition: Advancing Accuracy Through Xception and Bilinear Pooling Fusion

Authors: Hoang-Tu Vo, Nhon Nguyen Thien, Kheo Chau Mui

PAGE 1045 – 1051

Paper 114: Predicting Quality Medical Drug Data Towards Meaningful Data using Machine Learning

Authors: Suleyman Al-Showarah, Abubaker Al-Taie, Hamzeh Eyal Salman, Wael Alzaydat, Mohannad Alkhalileh

PAGE 1052 – 1059

Paper 115: Incorporating Learned Depth Perception Into Monocular Visual Odometry to Improve Scale Recovery

Authors: Hamza Mailka, Mohamed Abouzahir, Mustapha Ramzi

PAGE 1060 – 1068

Paper 116: Enhancing Precision in Lung Cancer Diagnosis Through Machine Learning Algorithms

Authors: Nasareenbanu Devihosur, Ravi Kumar M G

PAGE 1069 – 1077

Paper 117: Generating Nature-Resembling Tertiary Protein Structures with Advanced Generative Adversarial Networks (GANs)

Authors: Mena Nagy A. Khalaf, Taysir Hassan A Soliman, Sara Salah Mohamed

PAGE 1078 – 1088

Paper 118: Prediction of Heart Disease using an Ensemble Learning Approach

Authors: Ghalia A. Alshehri, Hajar M. Alharbi

PAGE 1089 – 1097

Paper 119: A Framework for Agriculture Plant Disease Prediction using Deep Learning Classifier

Authors: Mohammeld Baljon

PAGE 1098 – 1111

Paper 120: Lung Cancer Classification using Reinforcement Learning-based Ensemble Learning

Authors: Shengping Luo

PAGE 1112 – 1122

Paper 121: Secure Data Sharing in Smart Homes: An Efficient Approach Based on Local Differential Privacy and Randomized Responses

Authors: Amr T. A. Elsayed, Almohammady S. Alsharkawy, Mohamed S. Farag, S. E. Abo-Youssef

PAGE 1123 – 1132

Paper 122: The Implementation of Image Conceptualization Split-Screen Stitching and Positioning Technology in Film and Television Production

Authors: Zhouzhou Deng, Rongshen Zhu

PAGE 1133 – 1140

Paper 123: Rural Landscape Design Data Analysis Based on Multi-Media, Multi-Dimensional Information Based on a Decision Tree Learning Algorithm

Authors: Ning Leng, Hongxin Wang

PAGE 1141 – 1146

Paper 124: Intelligent Detection System for Electrical Equipment based on Deep Learning and Infrared Image Processing Technology

Authors: Mingxu Lu, Yuan Xie

PAGE 1147 – 1155

An Empirical Internet Protocol Network Intrusion Detection using Isolation Forest and One-Class Support Vector Machines

Gerard Shu Fuhnwi¹, Victoria Adedoyin², Janet O. Agbaje³

Gianforte School of Computing, Montana State University, Montana 59715, USA¹

Department of Chemistry, Montana State University, Montana 59715, USA²

Department of Mathematical Sciences, Montana Technological University, Montana 59701, USA³

Abstract—With the increasing reliance on web-based applications and services, network intrusion detection has become a critical aspect of maintaining the security and integrity of computer networks. This study empirically investigates internet protocol network intrusion detection using two machine learning techniques: Isolation Forest (IF) and One-Class Support Vector Machines (OC-SVM), combined with ANOVA F-test feature selection. This paper presents an empirical study comparing the effectiveness of two machine learning algorithms, Isolation Forest (IF) and One-Class Support Vector Machines (OC-SVM), with ANOVA F-test feature selection in detecting network intrusions using web services. The study used the NSL-KDD dataset, encompassing hypertext transfer protocol (HTTP), simple mail transfer protocol (SMTP), and file transfer protocol (FTP) web services attacks and normal traffic patterns, to comprehensively evaluate the algorithms. The performance of the algorithms is evaluated based on several metrics, such as the F1-score, detection rate (recall), precision, false alarm rate (FAR), and Area Under the Receiver Operating Characteristic (AUCROC) curve. Additionally, the study investigates the impact of different hyper-parameters on the performance of both algorithms. Our empirical results demonstrate that while both IF and OC-SVM exhibit high efficacy in detecting network intrusion attacks using web services of type HTTP, SMTP, and FTP, the One-Class Support Vector Machines outperform the Isolation Forest in terms of F1-score (SMTP), detection rate (HTTP, SMTP, and FTP), AUCROC, and a consistent low false alarm rate (HTTP). We used the t-test to determine that OCSVM statistically outperforms IF on DR and FAR.

Keywords—HTTP; SMTP; FTP; ANOVA F-test; AUCROC; OC-SVMs; FAR; DR; IF

I. INTRODUCTION

Network Intrusion can be referred to as an unauthorized penetration of a computer in an establishment or an address in one's assigned domain [1]. The nature and types of network intrusion have evolved over the years and become more rampant in recent years [2].

An intrusion can be passive or active. In passive intrusion, the penetration is gained stealthily and without detection, while in active intrusion, changes to network resources are affected. Intrusion can either come from an insider or an outsider. By insider, we mean an employee, customer, or business partner. Outsider means someone not connected to the organization. Network intrusions can occur in different ways. Some announce their presence by defacing the website, while others are malicious, with the goal of siphoning off data

until it's discovered. Some redirect users who are unaware of their website through cracking passwords or mimicking your website [1]. Sometimes, intruders absorb network resources intended for other uses or users, which can lead to a denial of service [3]. These unauthorized penetrations on the digital network are imperil on many occasions the security of networks and their data [4].

Network security breaches are rapidly increasing and result in a significant amount of loss to organizations, and often leads to a loss of confidence in them from their unaware customers that have fallen victims. The IBM report shows that the average cost of a data breach has risen 12 percent over the past five years to 3.92 million dollars per incident on average [5]. This is more than the cost of a breach caused by a system glitch or human error.

Many researchers have carried out research and projects on network intrusion detection [6], [7], [8]. Wang and Battiti identified intrusions in computer networks with principal component analysis [9]. Liao and Vemuri used a k-nearest neighbour classifier for intrusion detection [10]. Gaffney and Ulvila evaluated intrusion detectors using a decision theory approach [11]. But this area still longs for more work as a result of the rapid rise in network intrusion. Therefore, we need to design an efficient algorithm that can successfully defend against network intrusions in an ever-evolving threat landscape. To achieve proactive security control, organizations must put in place a good network security infrastructure and leverage the potential of machine learning, which has the capability of automatically and continuously detecting network intrusions. This will help block intruders and prevent them from achieving their goals. The remainder of this paper is organized as follows: In Section II, we briefly review some related work in anomaly detection based Network Intrusion Detection. Section III gives a description of the algorithms used in this paper. Section IV analyzes the empirical evaluation, where we review the data sets used, evaluation metrics description, results, and result discussion. Section V covers the conclusion.

II. RELATED WORK

Liu and Ting [23] focused on using an Isolation Forest to detect anomalies that have many applications in the areas of fraud detection, network intrusion, medical and public health, industrial damage detection, and so on. The goal here is to build a tree-based structure that isolates anomalies

rather than profiles anomalies like in the previous methods such as classification-based methods [12], and clustering-based methods [13]. Their proposed method, called Isolation Forest, builds a collection of individual tree structures that recursively partition a given data set, where anomalies are instances with a short path length on the trees. The anomaly score is used to determine instances that are anomalies, and has values between 1 and 0, with a score close to 1 being an anomaly and vice versa. The authors compared their results with other methods for anomaly detection techniques [14] like ORCA, LOF, and RF on real-world data sets with high dimensions and large data sizes using the metric AUC (Area Under the Curve) and run times. [15] proposed a hybrid of SVM and decision trees in classifying attacks of different forms of intrusion in knowledge discovery and data mining 1999 (KDDCUP99) data.

In [16], Sarumi et al. compared SVM and Apriori using Network Security Laboratory Knowledge Discovery and Data Mining (NSL-KDD) data and the University of South Wales NB 2015 (UNSW NB-15) dataset. From their results, they concluded that SVM outperformed Apriori in terms of accuracy, while Apriori showed a better performance in terms of speed.

In [17], Farnaaz and Jabbar proposed a detection intrusion system using random forest. Experimental results were conducted on the NSL-KDD dataset. Empirical results show that the proposed model achieved a low false alarm rate and a high recall. Similarly, [18], [19], [20], and [21] applied machine learning techniques for network intrusion detection systems.

All the above mentioned papers discuss intrusion detection methods without any statistics to compare their results, attacks using web services, and no user guidance for using the proposed algorithms. To overcome this, one can look at the statistical significance of the various evaluation metrics based on the different machine learning algorithms proposed by them and also change the various parameters in the machine learning algorithms to observe their performance.

This paper compares the performance of One-class SVM and Isolation Forest machine learning algorithms in network intrusion using a two-sample t-test and parameter alternation to provide some guidance on these algorithms' usage to new researchers in this field. Our approach can also guide evaluating and analyzing these techniques in solving intrusion detection problems. Also, this method can overcome one of the main challenges of intrusion detection techniques, accurate representative labels for normal and abnormal instances, which is a significant concern. To overcome this challenge in most intrusion detection problems, our approach can be used as a pre-labeling technique and then supervised anomaly detection techniques to solve intrusion detection problems. Overall, our empirical results demonstrate the potential of Isolation Forest and One-Class SVM and provide valuable insights for future research in this field.

III. METHODS

This section presents the intrusion detection approach used in this paper. These approaches include the ANOVA F-test, the Isolation Forest, the One-Class Support Vector Machines, and the two-sample t-test.

A. ANOVA F-test

The ANOVA F-test, or Analysis of Variance F-test, is a statistical technique used to compare the means of two or more groups to determine whether significant differences exist. It is commonly employed in feature selection or variable ranking tasks, where the goal is to identify the most relevant features or variables for a particular analysis or model.

Applying the ANOVA F-test to a dataset can rank features based on their F-statistic or p-value. Features with high F-statistic values or low p-values are considered more relevant, as they exhibit significant differences between the groups or classes. These relevant features can then be selected for further analysis or modeling, while less informative features can be discarded to reduce dimensionality and improve computational efficiency. In the case of web network intrusion detection, the ANOVA F-test can be used to identify the most discriminative features that differentiate between normal network traffic and malicious intrusion attempts. By selecting the most significant features, it is possible to improve the performance and efficiency of intrusion detection systems by focusing on the most relevant information and reducing noise or irrelevant variables.

B. Isolation Forest (IF)

IF has been applied in different scenarios. Isolation Forest is an unsupervised learning algorithm for anomaly detection that works on the principle of isolating anomalies, instead of the most common technique of profiling normal points [22] and [23]. It is different from other distance and density based algorithms (see Fig. 1). The underlying assumption for this algorithm is that fewer instances of anomalies result in a smaller number of partitions (shorter path length) and the instances with distinguishable attribute values are more likely to be separated in early partitioning [24]. This implies that data points that have a shorter path length are likely to be anomalies. The necessary input parameters for building Isolation Forest algorithm are the subsampling size, the number of trees, and the height of the tree [24]. The subsampling size was suggested to be smaller for the machine learning algorithm to function faster and yield a better detection result [25]. We can use log to base 2 (number of data points) to get the depth of trees needed, but the path length converge before $t = 100$. [25].

Algorithm 1: $iForest(X, t, \psi)$
Inputs: X – input data, t – number of trees, ψ – subsampling size
Output: a set of t $iTrees$
1: Initialize Forest
2: set height limit $l = \text{ceiling}(\log_2 \psi)$
3: for $i = 1$ to t do
4: $X' \leftarrow \text{sample}(X, \psi)$
5: $Forest \leftarrow Forest \cup iTrees(X', l)$
6: end for
7: return Forest

Fig. 1. Algorithm 1.

C. One-Class Support Vector Machines

One-Class Support Vector Machines (OC-SVMs) [26] are a natural extension of SVMs. One-Class SVM is an unsupervised learning technique capable of differentiating test samples from a particular class from other classes. The One-Class SVM works on the basics of minimizing the hypersphere of one

class in the training set and then considers every other class not within the hypersphere as anomalies or outliers. In order to identify suspicious observations, an OC-SVM estimates a distribution that encompasses most of the observations and then labels as “suspicious” those that lie far from it with respect to a suitable metric. This model uses different kernel functions or hyperspheres: linear, radial basis, sigmoid, and polynomial.

D. Two Sample t-test

The two-sample t-test, also known as the independent samples t-test or unpaired t-test, is a statistical hypothesis test used to compare the means of two independent groups to determine if there is a significant difference between them. The test assumes that the data is normally distributed and that the variances of the two groups are equal (although there are modifications available if this assumption does not hold). In order to compare the performance of IF and OCSVM with the ANOVA F-test, we used a two-sample t-test to test whether there is a significant difference between the mean performances of DR, FAR, F_1 score, AUCROC, and precision. The null hypothesis (H_0) for the two-sample t-test states that there is no significant difference between the mean performances of the two models, while the alternative hypothesis (H_1) states that there is a significant difference, unlikely to have occurred by chance, between the mean performances of the two models.

IV. EMPIRICAL EVALUATION

A. Data Description

The NSL–KDD dataset is an improved version of the KDD99 dataset, in which a large amount of data redundancy has been removed [27]. This dataset has the same attributes as the KDD99 having 41 features that are labeled as either normal or attacks using different web services (http, smtp, ftp, etc.). The NSL–KDD dataset repository has two files: KDDTrain.txt and KDDTest.txt. Table I shows the attack categories using different services and the number of data points per category in the NSL–KDD train and test datasets. The NSL–KDD dataset has 125973 data points in the training dataset and 22544 in the testing dataset.

TABLE I. THE ATTACK TYPES (CLASS) USING DIFFERENT INTERNET PROTOCOLS (HTTP, SMTP AND FTP), THE NUMBER OF RECORDS IN THE NSL-KDD TRAINING AND TESTING DATASET

Attacks using different Internet Protocol	No. of records		Attack Types (class)
	Training	Testing	
Normal	45,078	7,291	Normal traffic data
HTTP	2,289	1,180	Worm, Land, Smurf, Udpstorm, Teardrop, Pod, Mailbomb, Neptune, Process table, Apache2, Back
SMTP	284	316	Ipsweep, Nmap, Satan, Portsweep, Mscan, Saint, WarezClient, Worm, SnmpGetAttack, WarezMaster,
FTP	648	48	Imap, SnmpGuess, Named, MultiHop, Phf, SPy, Sendmail, Ftp_Write, Xsnoop, Xlock, Guess_Password

B. Data Pre-processing

The NSL-KDD dataset has 41 features, each representing an attack type described in Section 4.1 and an attack category (class feature). These features are both numeric (38 features) and categorical (3 features). The categorical features are protocol_type (3 types), service (70 types), and flag (11 types) that need to be converted to numeric features. We want to extract the most popular attacks caused by using different internet protocols (the service feature). These widespread attacks use web services (internet protocols) such as hypertext transfer protocol (HTTP), simple mail transfer protocol (SMTP), and file transfer protocol (FTP). After extracting the various internet protocols, the attack types (class) feature is labeled with a numeric type, starting with Normal, labeled as 0 and 1 for the different attack types.

Using ANOVA F-test feature elimination, the most relevant features with the highest F-statistic values in the dataset are identified, eliminating the least important features. These features are src bytes (number of data bytes transferred from source to destination in a single connection), dst bytes (number of data bytes transferred from destination to source in a single connection), and duration.

C. Confusion Matrix

The performance of machine learning techniques can be evaluated using different parameters. These parameters are calculated using True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN) as shown in the confusion matrix [28] in Table II. The following parameters are used to evaluate our proposed approach.

1) *Detection Rate (DR)*: It is the ratio between the total number of attacks detected by the NIDS and the total number of attacks present in the dataset [17] which can be calculated using the formula:

$$DR = \frac{TP}{TP + FN}$$

2) *Precision*: This measures the fraction of examples predicted as attacks that turned out to be attacks, which can be calculated using the formula:

$$Precision = \frac{TP}{TP + FP}$$

3) F_1 *Score*: It is the harmonic mean of the fraction of examples predicted as attacks that turned out to be attacks (precision). It can also be described as the ratio between the total number of attacks detected by the NIDS and the total number of attacks present in the dataset (the detection rate) which can be calculated using the formula:

$$F_1 \text{ Score} = \frac{2 * TP}{2 * TP + FN + FP}$$

4) *False Alarm Rate (FAR)*: It is the fraction of non attacks that are misclassified as attacks, which can be calculated using the formula:

$$FAR = \frac{FP}{FP + TN}$$

TABLE II. CONFUSION MATRIX: A CONTINGENCY CONTAINING FOUR METRICS, TRUE POSITIVE (TP), TRUE NEGATIVE (TN), FALSE POSITIVE (FP), AND FALSE NEGATIVE (FN).

Attack		Predicted Class	
		Yes	No
Actual class	Yes	TP	FN
	No	FP	TN

5) *Receiver Operating Characteristic (ROC) Curve*: The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of binary classification models in machine learning. It is created by plotting the ratio between the total number of attacks detected by the NIDS to the total number of attacks present in the dataset (detection rate) against the fraction of non-attacks that are misclassified as attacks (False Alarm Rate) at various classification threshold levels. The area under the curve (AUC) of the ROC quantifies the overall performance of the classification model. AUC values range from 0 to 1, with a value of 0.5 representing a random classifier and a value of 1 indicating a perfect classifier. A higher AUC value suggests a better-performing classification model.

A good NIDS should have high detection rates, precision, AUCROC, F_1 score but low FAR.

Most machine learning algorithms are evaluated using predictive accuracy, but this is not appropriate for network intrusion detection because it mostly involves imbalanced data. In terms of imbalanced data, we mean that the proportion of data points in each class is not approximately equal. The evaluation metrics adopted in this paper for evaluation and comparison of our models are standard AUC (Area under curve). The area under the receiver operating curve gives an average measure of performance across all possible classification thresholds.

D. Experimental Results

All experiments were performed in Python with alternating parameters for Isolation Forest (Sklearn) and One-class support vector machines (Sklearn) using the Intel(R) Core(TM) i7-10510U CPU at 1.80 GHz and 2.30 GHz processor with 16 GB of RAM. Training and testing of the Isolation Forest took four seconds, while it took 40 seconds to train the One-Class support vector machine model on the three selected features from the NSL–KDD dataset. The experimental results for One-class support vector machines and Isolation Forest on different performance metrics are shown in Table III and Table IV respectively.

E. Discussion of Results

In Table III, the polynomial kernel outperformed the other kernels on HTTP and SMTP subsets with high DR, F_1 Score, AUCROC, Precision, and low FAR. On the other hand, in Table III, the sigmoid kernel performed much better than the different kernels on the FTP subset.

In Table IV, isolation forest with 100 estimators on the HTTP subset achieved the highest DR, F_1 Score, AUCROC,

Precision, and low FAR. The SMTP and FTP subset performs best on the evaluation metrics with 50 estimators. Generally, both Isolation Forest and One-class support vector machines didn't perform well on the FTP subset, having very high FAR and low DR, F_1 Score, AUCROC, and Precision. It is evident in Table III and IV that the one-class support vector machines outperform Isolation Forest on all subsets, that is, HTTP, SMTP, and FTP having high DR, F_1 Score, and low FAR. Statistical analysis of overall performance on the one-class support vector machines and Isolation Forest results used a two-sample t-test with two-tailed probability to determine if each model's DR and FAR score on the test data yielded statistically significant differences ($p < 0.05$). In the HTTP, SMTP, and FTP, the one-class support machines had a significantly different DR, and FAR score ($p < 0.001$), which showed that our hypothesis was accepted.

Our approach improved detection capabilities on selected attack types when compared to other models and benchmarks in related work. The RFE process identified several key features highly relevant to network intrusion detection. These features align with our expectations and prior research [18], [20], and [25], confirming the importance of specific traffic characteristics in detecting malicious activities. The combination of IF, OC-SVM and ANOVA F-test not only improved the model's performance but also reduced the complexity of the model by eliminating redundant and irrelevant features.

The practical implications of our findings are significant for the field of network intrusion detection. The improved detection rates offered by our approach can help security practitioners identify and respond to cyber threats more effectively. Additionally, reducing false positives and negatives can minimize the operational overhead of manually investigating false alarms. Furthermore, our approach demonstrates potential scalability and adaptability for different network environments and evolving cyber threats.

V. CONCLUSION AND FUTURE WORK

The experiments performed on the NSL-KDD network intrusion data show that One-class support vector machines had the overall best performance in terms of DR and FAR scores over the Isolation Forest, with the best performance obtained by tuning the default parameters in both algorithms. Also, the number of estimators in Isolation Forest is comparable; using 100 and 50 estimators outperformed 200 estimators.

Therefore, One-class support is a good model for network intrusion detection by changing the default parameters in Sklearn. Also, polynomial or sigmoid kernel functions could be the best kernels to choose when using One-Class SVM on network intrusion data. Because of the usage of feature selection, the computational cost decreases (four seconds for Isolation Forest and forty seconds), and our experimental results indicate that our proposed approach increases the DR, F_1 score, AUCROC, and precision and decreases FAR for three types of attacks. We equally compared one-class support vector machines and Isolation Forest selected attack types using a two-sample t-test and found that our proposed approach (with fewer features) is promising. For future work, we will experiment with deep learning approaches like GANs and autoencoders since they are capable of handling data of higher

TABLE III. ONE-CLASS SVM PERFORMANCE MEASURE ON NSL–KDD TEST

Attacks using different Internet Protocol	Kernel	Gamma	DR	F ₁ Score	AUCROC	Precision	FAR
HTTP	Linear	0.00005	0.9802	0.9303	0.6308	0.8852	0.0198
	Sigmoid	0.00005	0.9969	0.9386	0.8867	0.6383	0.0031
	Polynomial	0.00005	0.9969	0.9385	0.6378	0.8866	0.0031
SMTP	Linear	0.00005	0.8398	0.7228	0.4468	0.6345	0.1602
	Sigmoid	0.00005	0.9806	0.7953	0.5156	0.6689	0.0194
	Polynomial	0.00005	0.9838	0.7969	0.5172	0.6696	0.0162
FTP	Linear	0.00005	0.3333	0.5000	0.6667	1.0000	0.6667
	Sigmoid	0.00005	0.5208	0.6848	0.7604	1.0000	0.4791
	Polynomial	0.00005	0.3333	0.5000	0.6667	1.0000	0.6667

TABLE IV. ISOLATION FOREST PERFORMANCE MEASURE ON NSL–KDD TEST

Attacks using different Internet Protocol	Estimators	maximum samples	DR	F ₁ Score	AUCROC	Precision	FAR
HTTP	50	256	0.9618	0.9563	0.8402	0.9508	0.0382
	100	256	0.9631	0.9570	0.8409	0.9509	0.0369
	200	256	0.9619	0.9563	0.8399	0.9507	0.0381
SMTP	50	256	0.9725	0.7846	0.6697	0.9725	0.0275
	100	256	0.9709	0.7838	0.6697	0.9709	0.0291
	200	256	0.9693	0.7835	0.6699	0.9693	0.0307
FTP	50	256	0.2917	0.3043	0.6225	0.3182	0.7083
	100	256	0.2083	0.2273	0.5809	0.2500	0.7916
	200	256	0.2708	0.2857	0.6121	0.3023	0.7292

dimensions and also evaluate one-class support vector machines and Isolation Forest using supervised learning methods like the random forest, Xgboost, and cost sensitive support vector machines.

REFERENCES

- [1] Michael West: Chapter 2 - Preventing System Intrusions: Network and System Security (Second Edition), pp.29–56. Syngress, Boston, 2014. doi:10.1016/B978-0-12-416689-9.00002-2.
- [2] Thomas M. Chen and Patrick J. Walsh: Chapter 3 - Guarding Against Network Intrusions. Network and System Security (Second Edition). pp.57–82. Syngress, Boston, 2014. doi:10.1016/B978-0-12-416689-9.00003-4.
- [3] Robert Moskowitz: Network Intrusion: Methods of Attack, 2014.
- [4] Isabell Gaylord: Network Intrusion: How to Detect and Prevent It. Reducing Risk, United states Cybersecurity Magazine, 2020. <https://www.uscybersecurity.net/network-intrusion/>.
- [5] David Bisson: How to Foil the 6 Stages of a Network Intrusion, Tripwire State of security news, 2019. <https://www.tripwire.com/state-of-security/security-data-protection/security-hardening/6-stages-of-network-intrusion-and-how-to-defend-against-them/>.
- [6] Kumar, Amit and Maurya, Harish Chandra and Misra, Rahul: A research paper on hybrid intrusion detection system, International Journal of Engineering and Advanced Technology (IJEAT), vol.2, no.4, pp.294–297, Citeseer, 2013.
- [7] Li Tian and Wang Jianwen: Research on Network Intrusion Detection System Based on Improved K-means Clustering Algorithm, International Forum on Computer Science-Technology and Applications, vol. 1, pp. 76–79, 2009. doi:10.1109/IFCSTA.2009.25.
- [8] Dikshant Gupta and Suhani Singhal and Shamita Malik and Archana Singh: Network intrusion detection system using various data mining technique, International Conference on Research Advances in Integrated Navigation Systems (RAINS), pp.1–6, 2016. doi:10.1109/RAINS.2016.7764418.
- [9] Wang, Wei and Battiti, Roberto: Identifying intrusions in computer networks based on principal component analysis, University of Trento, 2005.
- [10] Liao, Yihua and Vemuri, V Rao: Use of k-nearest neighbor classifier for intrusion detection, Computers & security, vol.21, no.5, pp.439–448, Elsevier, 2002.
- [11] Ulvila, Jacob W and Gaffney Jr, John E: Evaluation of intrusion detection systems, Journal of Research of the National Institute of Standards and Technology, vol.108, no. 6, pp.453, 2003.
- [12] Abe, Naoki and Zadrozny, Bianca and Langford, John: Outlier detection by active learning. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp.504–509, 2006.
- [13] He, Zengyou and Xu, Xiaofei and Deng, Shengchun: Discovering cluster-based local outliers, Pattern Recognition Letters, vol.24, no.9, pp.1641–1650, Elsevier, 2003.
- [14] Fuhni, Gerard Shu and Agbaje, Janet O and Oshinubi, Kayode and Peter, Olumuyiwa James: An Empirical Study on Anomaly Detection Using Density-based and Representative-based Clustering Algorithms, Journal of the Nigerian Society of Physical Sciences, pp.1364–1364, 2023.
- [15] Mulay, Snehal A and Deval, PR and Garje, GV: Intrusion detection system using support vector machine and decision tree, International Journal of Computer Applications, vol.3, no.3, pp.40–43, Citeseer, 2010.
- [16] Sarumi, Oluwafemi A and Adetunmbi, Adebayo O and Adetoye, Fadekemi A: Discovering computer networks intrusion using data analytics and machine intelligence, Scientific African, vol.9, pp.e00500, Elsevier, 2020.
- [17] Farnaaz, Nabila and Jabbar, MA: Random forest modeling for network intrusion detection system, Procedia Computer Science, Elsevier, pp.213–217, (2016).
- [18] WS, Jenif D Souza and Parvathavarthini, B: Machine learning based intrusion detection framework using recursive feature elimination method, 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), pp.1–4, IEEE, 2020.
- [19] Ingre, Bhupendra and Yadav, Anamika: Performance analysis of NSL-KDD dataset using ANN, 2015 international conference on signal processing and communication engineering systems, pp.92–96, IEEE, 2015.
- [20] Aljawarneh, Shadi and Aldwairi, Monther and Yassein, Muneer Bani: Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model, Journal of Computational Science, pp.152–160, Elsevier, 2018.

- [21] Hamed, Tarfa and Dara, Rozita and Kremer, Stefan C: Network intrusion detection system based on recursive feature addition and bigram technique, *Computers & Security*, pp.137–155, Elsevier, 2018.
- [22] Wikipedia contributors: Isolation forest — Wikipedia, The Free Encyclopedia, 2020, https://en.wikipedia.org/w/index.php?title=Isolation_forest&oldid=985700362.
- [23] Liu, Fei Tony and Ting, Kai Ming and Zhou, Zhi-Hua: Isolation-based anomaly detection, Eighth IEEE International Conference on Data Mining, pp.413–422, IEEE, 2008.
- [24] Arunraj, Nari S and Hable, Robert and Fernandes, Michael and Leidl, Karl and Heigl, Michael: Comparison of supervised, semi-supervised and unsupervised learning methods in network intrusion detection system (NIDS) application, *Anwendungen und Konzepte der Wirtschaftsinformatik*, no.6, 2017.
- [25] Liu, Fei Tony and Ting, Kai Ming and Zhou, Zhi-Hua: Isolation-based anomaly, *ACM Transactions on Knowledge Discovery from Data (TKDD)*. vol.6, no.1, pp.1–39, Acm New York, NY, USA, 2012.
- [26] Larry M. Manevitz and Malik Yousef: One-Class SVMs for Document Classification, *Journal of Machine Learning Research* 2, pp.139–154, 2001.
- [27] Ring, Markus and Wunderlich, Sarah and Scheuring, Deniz and Landes, Dieter and Hotho, Andreas, “A survey of network-based intrusion detection data sets,” *Computers & Security*, vol.86, pp. 147–167, Elsevier, 2019.
- [28] Beauxis-Aussalet, Emma and Hardman, Lynda, “IEEE Conference on Visual Analytics Science and Technology (VAST)-Poster Proceedings,” pp. 1–2, 2014.

Ensemble Security and Multi-Cloud Load Balancing for Data in Edge-based Computing Applications

Raghunadha Reddi Dornala
Cloud Architect, USA

Abstract—Edge computing has gained significant attention in recent years due to its ability to process data closer to the source, resulting in reduced latency and improved performance. However, ensuring data security and efficient data management in edge-based computing applications poses significant challenges. This paper proposes an ensemble security approach and a multi-cloud load-balancing strategy to address these challenges. The ensemble security approach leverages multiple security mechanisms, such as encryption, authentication, and intrusion detection systems, to provide a layered defense against potential threats. By combining these mechanisms, the system can detect and mitigate security breaches at various levels, ensuring the integrity and confidentiality of data in edge-based environments. The multi-cloud load balancing strategy also aims to optimize resource utilization and performance by distributing data processing tasks across multiple cloud service providers. This approach takes advantage of the flexibility and scalability offered by the cloud, allowing for dynamic workload allocation based on factors like network conditions and computational capabilities. To evaluate the effectiveness of the proposed approach, we conducted experiments using a realistic edge-based computing environment. The results demonstrate that the ensemble security approach effectively detects and prevents security threats, while the multi-cloud load balancing strategy with edge computing to improve the overall system performance and resource utilization.

Keywords—Edge computing; cloud computing; dynamic load balancing; fog computing; multi-cloud load balancing

I. INTRODUCTION

With the rapid advancement of edge computing and cloud technology, the aviation industry has witnessed significant transformations in flight management systems. These systems now rely on distributed edge computing infrastructure and leverage multiple cloud service providers for improved performance and reliability [1]. To ensure secure and efficient operations, ensemble security and multi-cloud load balancing techniques have become crucial in managing flight management cloud applications [2]. Ensemble security integrates multiple security measures and protocols to protect flight management systems from cyber threats. As edge computing extends the attack surface, incorporating diverse security mechanisms becomes essential [3]. It can include encryption, authentication, access control, intrusion detection systems, and advanced threat intelligence to safeguard critical flight data and systems [4].

In addition to security, effective load balancing across multiple cloud providers is paramount for flight management cloud applications [5]. Load balancing distributes incoming

network traffic across various cloud instances to optimize resource utilization, reduce latency, and enhance system scalability [6]. Multi-cloud load balancing ensures that flight management systems can efficiently utilize cloud resources, handle sudden traffic spikes, and deliver a seamless user experience. Ensemble security and multi-cloud load balancing are intricately linked in edge computing environments [7]. Organizations can dynamically distribute traffic based on security policies and performance metrics by integrating security measures with load-balancing algorithms [8]. It allows flight management cloud applications to balance the workload efficiently across different cloud providers while ensuring data confidentiality, integrity, and availability [9].

This study explores the challenges and opportunities associated with ensemble security and multi-cloud load balancing in the context of flight management cloud applications. We will discuss the essential requirements, potential solutions, and benefits of adopting these techniques in the aviation industry. Furthermore, we will examine real-world use cases and best practices for implementing ensemble security and multi-cloud load balancing to enhance the safety and performance of flight management systems in edge computing environments. By leveraging ensemble security and multi-cloud load balancing with the integration of edge computing with task scheduling, the aviation industry can fully utilize edge computing capabilities while maintaining robust security measures. It ensures flight management systems' safe and efficient operation and opens up new possibilities for advanced applications, such as real-time analytics, predictive maintenance, and intelligent decision-making.

The paper's organization is as follows: Section II presents the literature survey by explaining the existing models and their drawbacks. Section III explained the Intrusion Detection and Prevention System with Multiple Loads Balancing. Section IV introduced edge computing with cloud computing. Section V presented the parameters used to analyze the performance of the proposed approach. Section VI introduced the conclusion for the overall work.

II. LITERATURE SURVEY

Kuppusamy et al. [10] proposed a combined dynamic model that integrates the reinforced optimization approach to increase the scheduling performance in Fog computing. The experiments are conducted using the FogSim simulator to create the data and an available tool that manages energy efficiency by using the resources in fog computing. The proposed system mainly focused on addressing the schedule

with less cost and analyzing the CPU processing time and assigned memory. Simulation results show that the proposed model performance is increased by 12% to 15% in terms of CPU usage and 6% to 11% of less energy usage for solving job scheduling issues. Jango et al. [11] proposed the two-phase scheduling model consisting of a Bi-factor approach, which classifies the task based on end date and preference scheduling using the advanced Jellyfish Algorithm (ADS). The parameters are measured based on the speed, capacity, task size, complete tasks, and total VM used for resource provisioning fog combined cloud platform. The model was tested on real-time and average datasets considering the small and high workload based on the assigned task. The performance is measured using QoS parameters that reduce the cost and other metrics. K. Cao et al. [12] introduced the edge-based integrated model that deploys the cloud application in heterogeneous servers that estimate the response time from data centers. The proposed model can process both offline and online phases. An edge-based optimized model is executed offline, and the online mobility-based gaming system is developed to overcome the issues. Results show that the response time is improved by 48.57% from the base station. J. Al-Jaroodi et al. [13] proposed a distributed model that combined with cloud and fog-integrated sCPS, named PsCPS. The proposed integrates the multiple clouds, fogs, and sCPS subsystems to provide better services to face many challenges for combining. R. Deng et al. [14] proposed a fog-cloud integrated model that reduces power consumption and measures transmitting delay. The proposed model primarily focused on resolving the workload assignment issue by allocating work between the fog and cloud with minimal usage and artificial service delay. The problem is divided into three sub-issues assigned to sub-systems and solved. The simulation results show that fog computing can significantly improve cloud computing efficiency in terms of bandwidth and dissemination latency. M. Guo et al. [15] introduced the delay-based workload allocation (DBWA) that solves the issues based on energy efficacy and delay-undertake workload allocation issues in IoT-integrated systems. The proposed approach would be improved if the Lyapunov drift-plus-penalty theory was adopted. Finally, the proposed approach obtains better performance in terms of efficient allocation. G. Qu et al. [16] proposed the Deep Meta Reinforcement Learning-based Offloading (DMRO) algorithm that merges various parallel DNNs with Q-learning to make fine-tuned learning approach to solve the adaptive decision-making in the dynamic platform. The proposed approach improved the offloading by up to 18.1%. A. Yousafzai et al. [17] proposed an effective integration-based measurement offloading system. The proposed approach needs the program borders at edge servers that lately combine applications. The outcome saves 45.45% of runtime and 85.45% of energy consumption. The proposed system finally obtained the resource-exhaustive IoT application processing in mobile edge computing (MEC). V. Mohammadian et al. [18] discussed various issues in the cloud load balancing model—the existing approach aimed to find the under-loaded and overloaded nodes and balanced loads. Various issues are identified and solved by the proposed approach in cloud computing. The proposed model obtained better results in terms of fault detection and integrated the

simulation tools considered into the account. A. Pandita et al. [19] introduced several scheduling approaches in cloud computing regarding fault tolerance. Several advantages and disadvantages are discussed in the comparison. In cloud computing, the researchers firmly designed an effective fault tolerance system. E. Gures et al. [20] provide the dynamic load balancing approach to solve the generic issue in cloud computing. Various solutions are discussed to solve the load balancing issue. A comparative analysis uses various load-balancing models and analyzes the performance. M. T. Sandikkaya et al. [21] proposed a novel security model (NSM) to secure the PaaS Providers over malicious behavior based on neighbors. The NSM is not like an intrusion detection system, which focuses on processors' usage and challenging resource access. Finally, the NSM approach analyzes the malicious traffic among 100k requests. Several classifiers are used to classify the malicious threads and show better accuracy. O. Sohaib et al. [22] propose a novel 2-tuple fuzzy semantic group selection method that relies on a technology-organization-environment (TOE) system and employs an approach for ordering preference by resemblance to ideal solution (TOPSIS). The TOPSIS is used to help small-to-medium-sized operations assess and make decisions on online e-commerce.

III. INTRUSION DETECTION AND PREVENTION SYSTEM WITH MULTIPLE LOADS BALANCING

An Intrusion Detection and Prevention System (IDPS) mainly focused on detecting and preventing the unauthorized users or malicious activities within a computer network. The following steps involved in IDPS:

1) *Data collection*: The IDPS collects network traffic and system log data to analyze for potential threats. This can include network packets, log files, and system events.

2) *Traffic monitoring*: The IDPS monitors network traffic in real-time or analyzes stored data to identify suspicious patterns or anomalies. This can be done using several signature-based and anomaly detection.

3) *Signature-based detection (SBD)*: SBD is mainly compared with the network traffic over system-generated database events, known as attack signatures. If any similarity is identified, then the intrusion is a better one.

$$\text{Match} = \text{Compare}(\text{Signature}, \text{Network Traffic}) \quad (1)$$

4) *Anomaly detection*: Anomaly detection identifies deviations from normal patterns of network traffic or system behavior. Statistical analysis or machine learning algorithms are often used to detect anomalies.

$$\text{Anomaly} = \text{Compare}(\text{Statistics}, \text{Network Traffic}) \quad (2)$$

5) *Behavior-based detection*: Behavior-based detection establishes a baseline of normal behavior for the network or system and identifies deviations from that baseline. It can involve monitoring user activity, network connections, or system processes.

$$\text{Deviation} = \text{Compare}(\text{Baseline}, \text{Network}) \quad (3)$$

6) *Alert generation:* When a potential intrusion is detected, the IDPS generates an alert or notification to notify system administrators or security personnel. The alert may contain information about the detected threat, its severity, and recommended actions.

7) *Response and prevention:* Based on the severity of the threat, the IDPS may take automated actions to prevent or mitigate the intrusion. This can include blocking suspicious network traffic, terminating malicious processes, or initiating security measures to protect the network.

8) *Logging and reporting:* The IDPS maintains logs of detected threats, alerts, and responses for future analysis and reporting. These logs can help in understanding attack patterns, identifying vulnerabilities, and improving security measures.

It's important to note that the specific equations or algorithms used in each step may vary depending on the implementation and technology used in the IDPS. The equations mentioned above are generalized representations to illustrate the concept.

IV. EDGE COMPUTING WITH CLOUD COMPUTING

Edge computing and cloud computing are two distinct paradigms in computing that serve different purposes and offer unique advantages. However, combined, they can create a robust and efficient computing ecosystem. This section discussed how edge computing and cloud computing work together to enhance the overall computing experience. The cloud platform provides multiple services to users based on the usage of computing resources, data storage, and cloud application over the internet. All these applications and resources are deployed in cloud data centers provided by the cloud service provider (CSP). Users can leverage these resources on demand without needing local infrastructure, reducing costs and improving scalability. Fig. 1 explains the step-by-step process of proposed approach with functionalities.

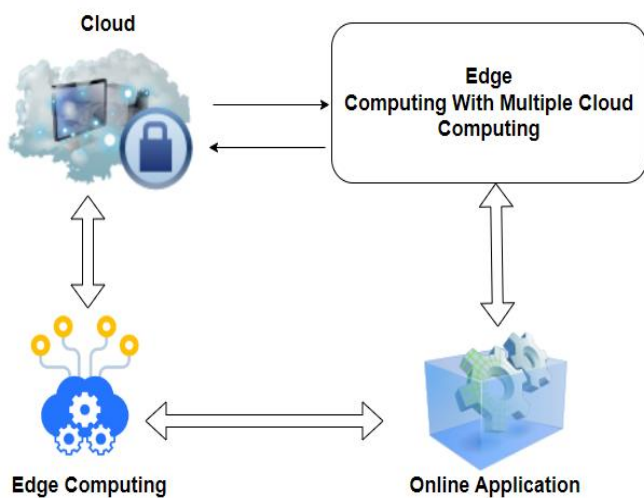


Fig. 1. Proposed system architecture.

On the other hand, edge computing brings computational capabilities closer to the data source or end-users. It involves performing data processing and analysis at the network's edge near the data source rather than sending all data to the cloud for processing. It reduces latency, minimizes bandwidth usage, and enables real-time decision-making, making it ideal for time-sensitive applications and services.

When edge computing and cloud computing are combined, they form a hybrid computing architecture that takes advantage of the strengths of both approaches. For example, sensors, smartphones, and IoT devices can perform initial data processing and filtering at the edge before selectively sending relevant data to the cloud for further analysis or storage. This method reduces the data sent to the cloud, saving bandwidth and allowing faster response times and more efficient resource utilization. The cloud is crucial in this hybrid architecture by providing a centralized and scalable infrastructure for advanced analytics, machine learning, and storage. The data from edge devices can be processed and analyzed in the cloud using sophisticated algorithms and models, taking advantage of the vast computing resources available. The insights and results can be returned to edge devices for immediate action or decision-making.

This edge and cloud computing combination is precious in various industries and applications. For example, edge computing can handle real-time sensor data processing in autonomous vehicles, while the cloud can perform high-level analytics and provide over-the-air updates. In healthcare, edge devices can collect patient data and perform initial analysis, while the cloud can store and process the data for long-term monitoring and research.

A. Load Balancing in Cloud Computing

Load balancing is an essential aspect of cloud computing that helps to optimize resource utilization, improve performance, and ensure the high availability of apps and services. Cloud computing is the delivery of computing resources such as virtual machines, storage, and applications via the internet. Load balancing distributes incoming network traffic across multiple servers or helps to avoid overloading and ensure efficient resource utilization. The primary goal of load balancing in cloud computing is to distribute workload evenly among servers or resources so that no single resource becomes overburdened while others remain underutilized. By evenly distributing the workload, load balancing helps to achieve optimal resource utilization, improve response times, and increase the overall throughput of the system.

Load balancing mechanisms in cloud computing can be categorized into two main types: static and dynamic load balancing. Static load balancing involves distributing the workload evenly based on predefined rules or static configurations. It is suitable for situations where the workload remains relatively constant and predictable. However, in dynamic and unpredictable environments, dynamic load balancing techniques are more effective. Dynamic load balancing utilizes various algorithms and techniques to monitor the system's workload continuously and adjust the distribution of requests accordingly. These techniques take into account factors such as server capacity, response time,

network latency, and current resource utilization to make intelligent decisions about workload distribution. Some popular dynamic load balancing algorithms include Round Robin (RR), Weighted Round Robin (WRR), Least Connections (LC), and Adaptive Load Balancing (ALB).

Load balancers also monitor the health and availability of servers and can automatically reroute traffic away from faulty or overloaded servers, ensuring high availability and fault tolerance. Cloud service providers often offer load balancing services as part of their cloud infrastructure offerings. These load balancers are typically provided as managed services, which means that the service provider takes care of the underlying infrastructure and maintenance, allowing users to focus on their applications and services.

B. Multi-Cloud Load Balancing Model

Round-robin (RR) and Consistent Hashing (CH) are two different algorithms used for load balancing in distributed systems. However, they can also be combined to achieve load balancing with Consistent Hashing.

Round-robin is a simple load balancing algorithm that distributes incoming requests evenly among a set of servers in a cyclic manner. Each request is assigned to the next available server in the rotation. This ensures that each server receives an equal number of requests over time.

For a given request i ($0 \leq i < M$), the server index is calculated as follows:

$$\text{server}_{\text{index}} = i \bmod N \quad (4)$$

Here, "mod" represents the modulo operation. The request is then sent to the server with the corresponding index.

Consistent Hashing is a more advanced load balancing algorithm that provides scalability and minimizes the redistribution of requests when a server is added or removed from the system. It uses a hash function to map each request to a specific server in the system based on the request's key or identifier.

To combine Round-robin with Consistent Hashing for load balancing, we can follow these steps:

Step 1: Initialize a list of servers.

Step 2: Assign each server a unique identifier or key.

Step 3: Create a hash ring to represent the distribution of keys across the servers.

Whenever a request comes in, apply the Consistent Hashing algorithm to determine the server responsible for serving that request.

Step 4: If the server is available and can handle the request, forward it to the server.

If the server is unavailable or overloaded, use Round-robin to select the next available server in the rotation and forward the request to that server.

Step 5: Continue the Round-robin rotation for subsequent requests until a server becomes available again.

Periodically update the hash ring and redistribute the keys when adding or removing servers from the system. By combining Round-robin with Consistent Hashing, we achieve both load balancing and the ability to scale the system dynamically without significant redistribution of requests. Round-robin ensures that the servers receive an equal share of requests, while Consistent Hashing minimizes the impact of adding or removing servers on the overall distribution of requests. This combined approach provides a balanced and efficient load balancing mechanism, ensuring optimal utilization of the server resources in a distributed system.

C. Task Scheduling for Multiple-Cloud

Organizations have increasingly turned to multi-cloud environments in recent years to meet their diverse computing needs. Multi-cloud refers to using multiple cloud service providers simultaneously, allowing businesses to leverage the strengths of different platforms and avoid vendor lock-in. However, managing and optimizing tasks across multiple clouds can be complex and challenging. Task scheduling is a critical aspect of managing cloud resources effectively. It involves determining when and where tasks or workloads should be executed to ensure efficient resource utilization and meet performance objectives. In multi-cloud environments, task scheduling becomes even more intricate as organizations must consider cost, performance, data locality, and compliance across multiple cloud providers. Multi-cloud task scheduling primarily aims to assign tasks to the most appropriate cloud resources based on various criteria, such as workload characteristics, resource availability, and user-defined policies. It involves making intelligent decisions to balance workload distribution, maximize resource utilization, minimize costs, and optimize performance across multiple clouds.

Several key challenges must be addressed to achieve effective task scheduling in a multi-cloud environment. These challenges include:

1) *Heterogeneity*: Each cloud provider offers different virtual machine types, pricing models, and service-level agreements. Task scheduling algorithms must account for this heterogeneity and make informed decisions about resource selection.

2) *Interoperability*: Ensuring seamless communication and data transfer between cloud providers is crucial. Task scheduling mechanisms should consider data locality and network latency to minimize data transfer costs and enhance performance.

3) *Scalability*: Multi-cloud environments often involve a large number of tasks and resources. Task scheduling algorithms must scale efficiently to handle the increased complexity and volume of scheduling decisions.

4) *Dynamicity*: Cloud environments are dynamic, with fluctuating workloads and resource availability. Task scheduling mechanisms should be adaptable and able to handle workload variations and changes in resource availability in real time.

5) *Cost optimization*: Multi-cloud environments introduce cost considerations due to varying pricing models and resource utilization. Task scheduling algorithms should minimize costs while meeting performance objectives and user-defined policies.

Addressing these challenges requires the development of intelligent task-scheduling algorithms and frameworks specifically designed for multi-cloud environments. These algorithms should consider workload characteristics, resource availability, performance requirements, and cost constraints to make optimal scheduling decisions.

D. Ensemble Security and Multi-Cloud Load Balancing

Ensemble Security and Multi-Cloud Load Balancing are two important concepts in the field of cloud computing and network security.

E. Ensemble Security

Ensemble Security refers to a comprehensive approach to securing computer networks and systems by using a combination of security measures and tools. The term "ensemble" implies the integration and coordination of various security components to create a stronger and more effective security system.

In ensemble security, multiple security mechanisms are employed to protect against a wide range of threats and vulnerabilities. This paper mainly used the mechanisms which includes intrusion detection and prevention systems (IDPS), data encryption. By combining these different security measures, organizations can create a layered defense system that provides multiple barriers against potential attacks.

The advantage of ensemble security is that it offers defense in depth, meaning that if one security measure fails, others can still provide protection. Additionally, ensemble security enables organizations to leverage the strengths of different security tools and technologies, thereby enhancing overall network security posture.

F. Multi-Cloud Load Balancing (MCLB)

In cloud computing, MCLB is a technique that distributes incoming network traffic across multiple cloud service providers (CSPs) or regions within a single CSP. Load balancing improves application performance, optimizes resource utilization, and ensures cloud-based services' high availability and scalability.

In a multi-cloud environment, organizations may choose to deploy their applications and services across different CSPs to take advantage of each provider's unique capabilities and to mitigate the risk of vendor lock-in. Multi-cloud load balancing allows organizations to distribute the incoming traffic across these different cloud environments, ensuring that the workload is evenly distributed and that no single cloud provider becomes overloaded.

Load balancers serve as go-betweens for users or clients and cloud resources. Incoming requests are intelligently distributed based on factors such as server capacity, latency in the network, and specific to application requirements. This helps in optimizing resource utilization, preventing

bottlenecks, and ensuring that the applications remain accessible and responsive. By implementing multi-cloud load balancing, organizations can achieve higher fault tolerance, improved performance, and scalability across multiple cloud environments. It also provides flexibility in managing and scaling resources based on changing demands and enables efficient utilization of cloud resources.

V. EXPERIMENTAL RESULTS

The experiments are conducted on flight management applications which can take Lakhs of requests from the user and managed by the proposed multi-cloud management system with high security. The total requests are analyzed by the proposed model is 10k, 20k and 30k requests. The performance metrics are given below:

1) *Response time (RT)*: It measures the time taken to process a request and provide a response to the client.

$$RT = \text{Complete time} - \text{request time} \quad (5)$$

2) *Throughput (TP)*: It represents the number of requests processed per unit of time.

$$TP = \text{Total} \frac{\text{Requests}}{\text{Time}} \quad (6)$$

3) *Utilization (UTL)*: It indicates the percentage of resources being used by a server or a load balancer.

$$UTL = \left(\frac{\text{Resource used}}{\text{Total Resource capacity}} \right) * 100 \quad (7)$$

4) *Error rate (ER)*: It measures the percentage of failed requests or the number of requests resulting in errors.

$$ER = \left(\frac{\text{Number of failed requests}}{\text{Total number of requests}} \right) * 100 \quad (8)$$

TABLE I. ANALYSIS OF PERFORMANCE METRICS FOR LIST OF ALGORITHMS FOR 10K REQUESTS FOR 10K FILES

Algorithms	RT (Sec)	TP (request/sec)	UTL (%)	ER (%)
Least Response Time (LRT)	56.45	67	56	32
IP Hash	51.23	76	61	28
MCLB	45.67	81	67	21

Table I shows the performance of several existing and proposed algorithms based on the given metrics. These metrics mainly shows the huge impact on final outcomes. The proposed model MCLB mainly focused on managing the load balancing among the servers and providing the security for the processing data. For every one request one file is requested by the users. The encryption and decryption is implemented at the data owners end to provide the high end security for the data. The MCLB improved with the integration of Edge computing obtained the performance in terms of given metrics. The total requests processed by the MCLB are 10k requests for 10 files. Fig. 2 shows the performance of algorithm based on the given parameters.

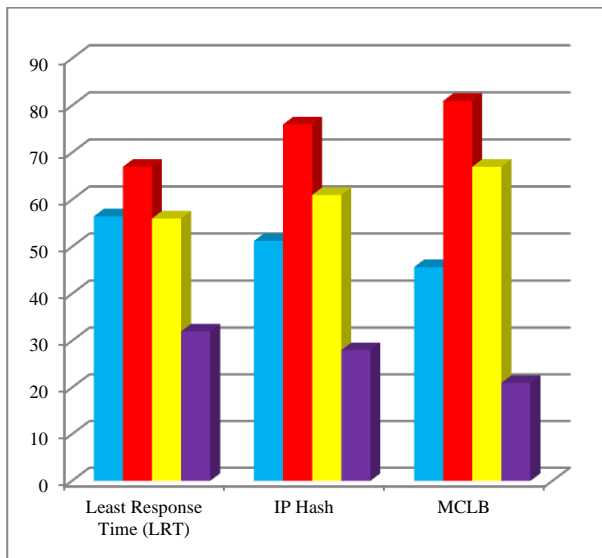


Fig. 2. Performance metrics for list of algorithms for 10k requests.

TABLE II. ANALYSIS OF PERFORMANCE METRICS FOR LIST OF ALGORITHMS FOR 20K REQUESTS

Algorithms	RT (Sec)	TP (request/sec)	UTL (%)	ER (%)
Least Response Time (LRT)	66.45	76	63.4	34.5
IP Hash	61.23	83	65.2	31.2
MCLB	65.67	85.5	69.34	22.2

Table II shows the performance of list of algorithms based on several parameters such as response time (RT) which is high for LRT and low for MCLB. The throughput is low for LRT and high for MCLB which is more efficient. The utilization of resources is high for MCLB and low for LRT. Finally the error rate (ER) shows the low rate for MCLB and high for LRT. Fig. 3 shows the overall performance for 20k requests.

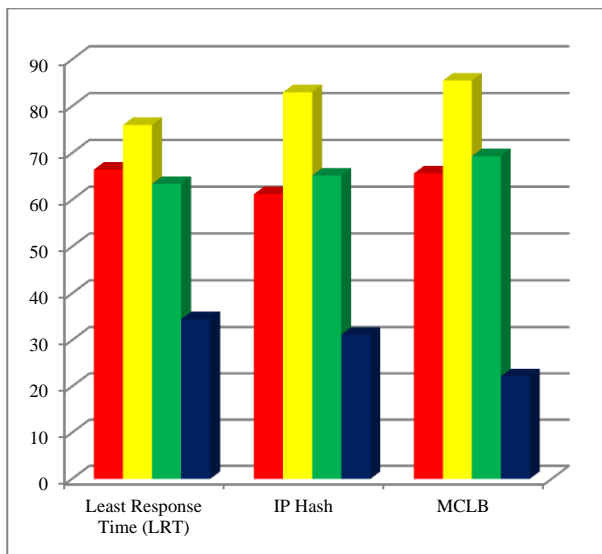


Fig. 3. Performance metrics for list of algorithms for 20k requests.

TABLE III. ANALYSIS OF PERFORMANCE METRICS FOR LIST OF ALGORITHMS FOR 30K REQUESTS

Algorithms	RT (Sec)	TP (request/sec)	UTL (%)	ER (%)
Least Response Time (LRT)	73.55	78.8	64.4	31.5
IP Hash	66.23	84.1	66.2	26.2
MCLB	68.67	86.3	70.34	18.2

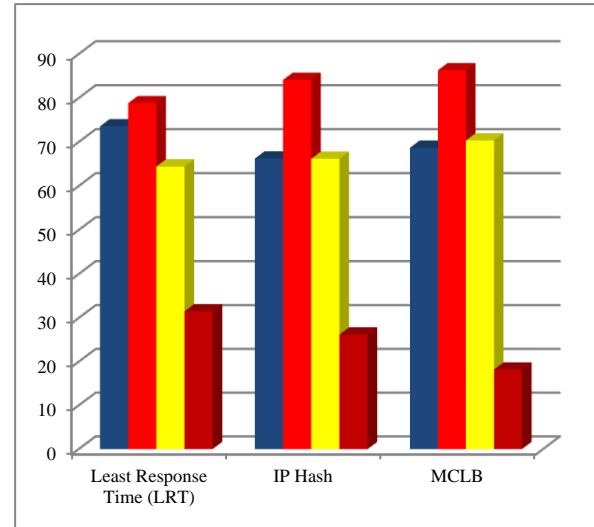


Fig. 4. Performance metrics for list of algorithms for 30k requests.

Table III shows the comparison between LRT, IP Hash and MCLB. The existing model LRT shows the low performance in terms of high RT (Sec), low TP and utilization of resources and high error rate. The proposed approach MCLB shows the better performance in terms of parameters given in Table III. Fig. 4 shows the overall performances for all the algorithms for 30k requests sent by various users present in VMs.

VI. CONCLUSION

Modern cloud computing and network security rely heavily on ensemble security and multi-cloud load balancing. Ensemble security combines multiple security measures to form a robust defense system, whereas multi-cloud load balancing optimizes resource distribution while improving application performance and availability in various cloud environments. The combination of edge computing and cloud computing provides a powerful computing model that takes advantage of the advantages of both paradigms. It allows efficient data processing, lowers latency, increases scalability, and enables real-time decision-making. Organizations can create a robust and flexible computing ecosystem that meets the diverse needs of today's rapidly evolving technological landscape by combining edge computing and cloud computing.

REFERENCES

- [1] Zhang, Wei-Zhe & Elgendy, Ibrahim & Hammad, Mohamed & Ilyasu, Abdullah & Du, Xiaojiang & Guizani, Mohsen & Abd El-Latif, Ahmed. (2021). Secure and Optimized Load Balancing for Multi-Tier IoT and Edge-Cloud Computing Systems. IEEE Internet of Things Journal. 1-1. 10.1109/IJOT.2020.3042433.

- [2] M. J. Priya and G. Yamuna, "Privacy preserving Data security model for Cloud Computing Technology," 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2022, pp. 1-5, doi: 10.1109/ICSTSN53084.2022.9761350.
- [3] X. Wei and Y. Wang, "Popularity-Based Data Placement With Load Balancing in Edge Computing," in IEEE Transactions on Cloud Computing, vol. 11, no. 1, pp. 397-411, 1 Jan.-March 2023, doi: 10.1109/TCC.2021.3096467.
- [4] A. Kishor, R. Niyogi, A. T. Chronopoulos and A. Y. Zomaya, "Latency and Energy-Aware Load Balancing in Cloud Data Centers: A Bargaining Game Based Approach," in IEEE Transactions on Cloud Computing, vol. 11, no. 1, pp. 927-941, 1 Jan.-March 2023, doi: 10.1109/TCC.2021.3121481.
- [5] W. Li, Q. Fan, W. Cui, F. Dang, X. Zhang and C. Dai, "Dynamic Virtual Machine Consolidation Algorithm Based on Balancing Energy Consumption and Quality of Service," in IEEE Access, vol. 10, pp. 80958-80975, 2022, doi: 10.1109/ACCESS.2022.3194514.
- [6] Wazir Zada Khan, Ejaz Ahmed, Saqib Hakak, Ibrar Yaqoob, Arif Ahmed, Edge computing: A survey, Future Generation Computer Systems, Volume 97, 2019.
- [7] Sulieman, N.A.; Ricciardi Celsi, L.; Li, W.; Zomaya, A.; Villari, M. Edge-Oriented Computing: A Survey on Research and Use Cases. Energies 2022, 15, 452. <https://doi.org/10.3390/en15020452>
- [8] K. Cao, S. Hu, Y. Shi, A. W. Colombo, S. Karnouskos and X. Li, "A Survey on Edge and Edge-Cloud Computing Assisted Cyber-Physical Systems," in IEEE Transactions on Industrial Informatics, vol. 17, no. 11, pp. 7806-7819, Nov. 2021, doi: 10.1109/TII.2021.3073066.
- [9] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge computing: Vision and challenges", IEEE Internet Things J., vol. 3, no. 5, pp. 637-646, Oct. 2016.
- [10] Kuppasamy, P., Kumari, N.M.J., Alghamdi, W.Y. et al. Job scheduling problem in fog-cloud-based environment using reinforced social spider optimization. J Cloud Comp 11, 99 (2022).
- [11] Jangu, N., Raza, Z. Improved Jellyfish Algorithm-based multi-aspect task scheduling model for IoT tasks over fog integrated cloud environment. J Cloud Comp 11, 98 (2022).
- [12] K. Cao, L. Li, Y. Cui, T. Wei and S. Hu, "Exploring placement of heterogeneous edge servers for response time minimization in mobile edge-cloud computing", IEEE Trans. Ind. Informat., vol. 17, no. 1, pp. 494-503, Jan. 2021.
- [13] J. Al-Jaroodi and N. Mohamed, "PsCPS: A distributed platform for cloud and fog integrated smart cyber-physical systems", IEEE Access, vol. 6, pp. 41432-41449, 2018.
- [14] R. Deng, R. Lu, C. Lai, T. Luan and H. Liang, "Optimal workload allocation in fog-cloud computing towards balanced delay and power consumption", IEEE Internet Things J., vol. 3, no. 6, pp. 1171-1181, Dec. 2016.
- [15] M. Guo, L. Li and Q. Guan, "Energy-efficient and delay-guaranteed workload allocation in IoT-edge-cloud computing systems", IEEE Access, vol. 7, pp. 78685-78697, 2019.
- [16] G. Qu, H. Wu, R. Li, and P. Jiao, "DMRO: A deep meta reinforcement learning-based task offloading framework for edge-cloud computing," IEEE Trans. Netw. Service Manage., vol. 18, no. 3, pp. 3448-3459, Sep. 2021.
- [17] A. Yousafzai, I. Yaqoob, M. Imran, A. Gani, and R. M. Noor, "Process migration-based computational offloading framework for IoT-supported mobile edge/cloud computing," IEEE Internet Things J., vol. 7, no. 5, pp. 4171-4182, May 2020, doi: 10.1109/JIOT.2019.2943176.
- [18] V. Mohammadian, N. J. Navimipour, M. Hosseinzadeh and A. Darwesh, "Fault-Tolerant Load Balancing in Cloud Computing: A Systematic Literature Review," in IEEE Access, vol. 10, pp. 12714-12731, 2022, doi: 10.1109/ACCESS.2021.3139730.
- [19] A. Pandita, P. K. Upadhyay and N. Joshi, "Fault Tolerance Based Comparative Analysis of Scheduling Algorithms in Cloud Computing," 2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET), Kottayam, India, 2018, pp. 1-6, doi: 10.1109/ICCSDET.2018.8821216.
- [20] E. Gures, I. Shayea, M. Ergen, M. H. Azmi and A. A. El-Saleh, "Machine Learning-Based Load Balancing Algorithms in Future Heterogeneous Networks: A Survey," in IEEE Access, vol. 10, pp. 37689-37717, 2022, doi: 10.1109/ACCESS.2022.3161511.
- [21] M. T. Sandikkaya, Y. Yaslan, and C. D. Ozdemir, "DeMETER in clouds: Detection of malicious external thread execution in runtime with machine learning in PaaS clouds," Cluster Comput., vol. 23, pp. 2565-2578, Dec. 2019.
- [22] O. Sohaib, M. Naderpour, W. Hussain, and L. Martinez, "Cloud computing model selection for e-commerce enterprises using a new 2-tuple fuzzy linguistic decision-making method," Comput. Ind. Eng., vol. 132, pp. 47-58, Jun. 2019.

Converting Data for Spiking Neural Network Training

Erik Sadovsky, Maros Jakubec, Roman Jarina

Department of Multimedia and Information-Communication Technologies-FEIT,
University of Zilina, Zilina, Slovak Republic

Abstract—The application of spiking neural networks (SNNs) for processing visual and auditory data necessitate the conversion of traditional neural network datasets into a format suitable for spike-based computations. Existing datasets designed for conventional neural networks are incompatible with SNNs due to their reliance on spike timing and specific preprocessing requirements. This paper introduces a comprehensive pipeline that enables the conversion of common datasets into rate-coded spikes, meeting processing demands of SNNs. The proposed solution is evaluated on Spike-CNN trained on Time-to-First-Spike encoded MNIST and compared with the similar system trained on the neuromorphic dataset (N-MNIST). Both systems have comparative precision; however the proposed solution is more energy efficient than the system based on neuromorphic computing. Since, the proposed solution is not limited to any specific data form and can be applied to various types of audio/visual content. By providing a means to adapt existing datasets, this research facilitates the exploration and advancement of SNNs across different domains.

Keywords—SNN; rate coding; spike timing; data conversion; MNIST

I. INTRODUCTION

Spiking Neural Networks (SNNs) have emerged as a highly promising research direction, bridging the gap between neuroscience and machine learning. By emulating the behaviour of biological neurons and their asynchronous communication through discrete spikes, SNNs offer a compelling computational framework for modelling and understanding neural processes [1], [2]. However, a significant obstacle in fully realizing the potential of SNNs lies in the lack of dedicated databases specifically designed for their training and evaluation.

In contrast to conventional neural networks that are trained using readily and widely available datasets such as MNIST or ImageNet (which are widely used in computer vision and pattern recognition fields among many others), SNNs require special data representations that capture the temporal dynamics of neural processing in a form of spikes. The precise timing of the spikes becomes crucial for encoding and processing information, necessitating a departure from traditional data formats [3]. Consequently, substantial research efforts are currently focused on developing comprehensive databases tailored explicitly for SNN training and evaluation.

Despite notable progress in the field of SNNs, the development of specialized databases for training and testing remains an ongoing research challenge. Currently, only a

limited number of publicly available datasets, such as the Spiking Neural Network Architecture (SNA), N-MNIST [4], DVS Gesture [5], and N-TIDIGITS [6], have been specifically designed for SNNs. These datasets enable training and testing of SNNs across various tasks, including decoding neural activity, image classification, gesture recognition, and speech recognition.

The availability of dedicated datasets is crucial for advancing the field of SNNs, as they serve as the foundation for training and evaluating network performance. However, existing datasets for SNNs (usually recorded using a specialized neuromorphic device) are limited in size and diversity, hindering the exploration of SNN capabilities across different domains and applications. This limitation underscores the pressing need to develop methodologies for converting and adapting conventional datasets into formats suitable for SNN training.

Our objective is to explain a processing pipeline for converting amplitude-based data into time/rate encoded spikes and compare a performance of SNN trained on such data with the performance of the SNN trained on specialized neuromorphic dataset. Such conversion process involves the transformation of input data into spike-based representations that preserve the temporal information necessary for accurate neural computation. This conversion necessitates careful consideration of various factors, including spike encoding schemes, spike rates, and the representation of spike timing. Moreover, it is essential to ensure that the converted data maintains the underlying structure and statistical properties of the original data to guarantee meaningful and reliable training of SNNs. By bridging the gap between conventional data formats and SNNs, our work aims to empower researchers and practitioners to overcome the limitations imposed by the scarcity of SNN-specific databases.

In this paper, we propose specifically an approach to converting image data (conventionally represented by pixel intensity values in matrix form) into spike-based representations incorporating temporal encoding techniques. We study and evaluate the proposed approach on MNIST dataset. To prove our concept, we compare the performance of developed SNN architectures trained on the spike-converted MNIST database with ones trained using the N-MNIST database, created by capturing MNIST images using neuromorphic Dynamic Vision Sensor (DVS) camera [7]. N-MNIST is a widely used benchmark dataset for evaluating the performance of SNN models in various tasks.

This paper is organized as follows. Section II introduces Spiking Neural Networks (SNNs) and emphasizes the need for new databases. Section III provides an overview of SNNs applications developed on the N-MNIST database. Section IV outlines the proposed methodology, including data conversion, database description, and network settings. Section V presents the results, demonstrating the effectiveness of SNNs trained on the newly acquired database. Finally, Section VI concludes the paper by summarizing the findings, discussing research implications and limitations, and providing recommendations for future studies.

II. SPIKING NEURAL NETWORK – THEORETICAL BACKGROUND

Despite the enormous efforts of scientists and evidently great progress in information and cognitive science, the human brain, with its billions of interconnected neurons, remains an enigmatic engine capable of complex cognitive processes. In recent years, there has been a surge of interest in developing computational models that emulate the functionality of biological neural networks. While traditional Artificial Neural Networks (ANNs) have been successful in numerous applications in diverse domains, they fall short in capturing the temporal dynamics and binary nature of spiking neurons observed in biological systems.

Spiking neural networks present a promising alternative to ANNs, providing a more biologically realistic approach to modelling neural computation. By communicating through discrete binary events known as spikes, SNNs mimic the action potentials observed in real neurons. This temporal coding scheme enables SNNs to capture the dynamics and synchronization observed in biological neural systems, opening new avenues for understanding brain function and developing advanced cognitive computing systems.

A. Neuronal Dynamics in SNNs

The core of a SNN lies in the dynamics of its constituent spiking neurons. Unlike traditional ANNs, which operate using real-valued activations, SNNs leverage the binary nature of spiking neurons to encode and process information through time.

1) *Integrate and fire model*: The Integrate and Fire (IF) [8] model represents a fundamental building block of SNNs. In this simplified model, the membrane potential of a neuron, denoted as V , integrates the input spike trains it receives. Once the membrane potential surpasses a threshold voltage, the neuron generates an output spike. The dynamics of the membrane potential in the IF model can be described as:

$$\frac{du}{dt} = R \cdot I(t), \quad u < V_{th} \quad (1)$$

where u denotes the membrane potential, the derivative du/dt represents the rate of change of the membrane potential, R is the membrane resistance, $I(t)$ represent the input spike train, and V_{th} is the threshold voltage.

2) *Leaky integrate and fire model*: The Leaky Integrate and Fire (LIF) model builds upon the IF model by incorporating the concept of leakage. In biological neurons,

the membrane potential gradually decays towards a resting potential due to ion leakage. The LIF model accounts for this phenomenon by including a leakage term in the dynamics of the membrane potential. The LIF model can be expressed as:

$$\tau_m \frac{du}{dt} = -(u - u_{rest}) + R \cdot I(t), \quad u < V_{th} \quad (2)$$

where τ_m represents the membrane time constant, u_{rest} is the resting potential. Fig. 1 provides a visual representation of the key elements and parameters characterizing a LIF neuron.

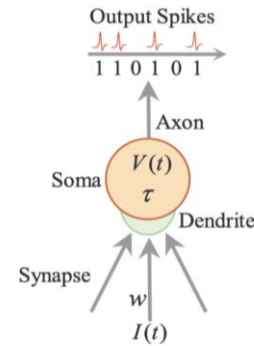


Fig. 1. LIF neuron characterized by membrane potential V , membrane time constant τ , input $I(t)$, and synaptic weight w .

B. Impulse Coding

Impulse coding is a fundamental aspect of SNNs, as it involves the transformation of data into an impulse-based format that enables efficient processing within these networks. The objective of impulse coding is to preserve relevant information while generating a stream of spikes. However, determining the importance of specific information and developing a unified approach to impulse coding remains a complex and context-dependent challenge. The following subsections briefly discuss three mechanisms of impulse coding in SNNs, namely rate encoding, temporal encoding, and population coding.

1) *Rate encoding*: Rate encoding, also known as rate coding, is a widely studied and utilized method of encoding information in SNNs. It's based on the assumption that the average firing rate of neurons over a specific time interval carries the desired information. By modulating the firing rate of neurons, different stimuli can be represented. The rate encoding approach offers a straightforward and intuitive method for representing information using spikes.

The strength of the stimulus representation is believed to increase with the firing rate. However, the precise mapping between firing rate and stimulus intensity can vary depending on the neural population and the specific encoding scheme employed. The rate encoding method provides a reliable means of representing information in SNNs and has been successfully applied in various applications.

2) *Temporal encoding*: Temporal encoding is another mechanism employed in impulse coding, which focuses on the precise timing of spikes to represent information. Instead of relying solely on the firing rate, temporal encoding

emphasizes the temporal order and precise timing of individual spikes. The relative timing of spikes across multiple neurons can convey specific features or patterns of stimuli.

In temporal encoding, the timing of spikes within a spike train carries the information, such as the duration between spikes or the occurrence of specific spike patterns. The brain has the remarkable ability to decode and interpret these temporal patterns to extract meaningful information. Temporal coding offers a rich representation that captures fine-grained details of stimuli and enables precise temporal processing in neural networks.

3) *Population encoding*: In addition to rate encoding and temporal encoding, population encoding has been introduced as a third category of impulse coding. Population encoding involves the joint activity of multiple neurons to encode information. Rather than relying on the individual firing rates or precise timing of spikes, population encoding considers the collective behaviour of a group of neurons.

The underlying principle of population coding is that the combined activity of a population of neurons carries information that cannot be represented by individual neurons alone. By analysing the distributed patterns of activity across the population, specific features or stimuli can be decoded. Population coding provides a powerful mechanism for encoding complex information and has been observed in various biological sensory systems.

III. RELATED WORK

In recent years, several studies have investigated a variety of SNN approaches, using the N-MNIST database, a widely used benchmark dataset for performance evaluation (see Section IV for details on N-MNIST). He et al. [9] compared the performance of the feedforward SNNs and recurrent neural networks (RNNs). The authors modified the N-MNIST database by compressing individual spiking events along the temporal axis and utilized the leaky integrate and fire (LIF) model as the spiking neuron model. Their findings indicated that SNNs generally outperformed conventional RNNs in terms of accuracy. However, with the adaptation of loss functions and the incorporation of Long Short-Term Memory (LSTM) networks, RNNs achieved competitive accuracy with SNNs. This study highlighted the advantage of SNNs in processing features represented by a sparse set of spikes. Another approach by Cohen et al. [10] introduced the method of inverse synaptic kernels for training SNNs on N-MNIST. The authors constructed a spiking neural network with a hidden layer comprising up to 10,000 neurons and achieved a high classification rate of 92.87% on the N-MNIST test subset. This work demonstrated the potential of leveraging biologically inspired principles to further enhance the performance of SNNs.

Wu et al. [11] proposed a novel architecture by incorporating a population of neurons in the output layer of a convolutional SNN. The output spike sequence from this layer represented population coding, which improved the discriminative capabilities of the network. The experiments conducted on the N-MNIST and DVS-CIFAR10 databases showed remarkable accuracies of 99.53% and 60.5%,

respectively. This study highlighted the effectiveness of population coding in visual recognition tasks using SNNs.

For event-based features as (SNN input data), Ramesh et al. [12] introduced the Event-Based structural Descriptor (EBD) that captures a spatio-temporal structure using a log-polar grid and applied it to various computer vision problems, including N-MNIST classification. They proved the efficacy of event-based representations in capturing spatiotemporal information and leveraging it for robust classification in SNN frameworks. Their classifier achieved a high accuracy of 97.95% on the N-MNIST test subset.

Addressing the challenges associated with SNN training, Shrestha et al. [13] explored the non-differentiability of the spike generation function and proposed a solution for converting existing databases into spike-based representations. They introduced the SLAYER algorithm, inspired by backpropagation, enabling the training of both feedforward and convolutional SNNs. The N-MNIST database was used to demonstrate the algorithm's effectiveness and the conversion process from image-based to spike-based representations.

Table I summarizes the accuracy achieved by various approaches on the N-MNIST dataset, providing a comprehensive performance comparison of the state-of-the-art SNN-based methods.

TABLE I. PERFORMANCE COMPARISON OF THE STATE-OF-THE-ART METHODS ON THE N-MNIST DATASET

Authors	Method	Accuracy (%)
Sironi et al., 2018 [14]	HATS	99.10
Lee et al., 2020 [15]	Spike based supervised gradient descent	99.09
Bi et al., 2019 [16]	Graph based object classification	99.00
Jin et al., 2019 [17]	HM2-BP	98.84
Yousefzadeh et al., 2018 [18]	Active perception with DVS	98.80
Wu et al., 2018 [19]	Spatiotemporal backpropagation	98.78
Lee et al., 2016 [20]	Training SNN using backpropagation	98.74
Ramesh et al., 2017 [12]	Event-Based Descriptor	97.95
Liu et al., 2020 [21]	Segmented probability-maximization	96.30
Kaiser et al., 2020 [22]	DECOLLE	96.00
Cohen et al., 2016 [10]	Inverse synaptic kernels for training SNN	92.87

The Spiking Heidelberg Digits (SHD) and the Spiking Speech Command (SSC) [23] datasets are both audio-based classification datasets that provide input spikes and output labels for different spoken digits and commands. Both of these datasets were created using a software conversion that was based on mathematical models of inner auditory system.

The IBM gestures dataset [24] contains spike trains of gesture movement recordings under different illumination conditions. It is one of the most popular real world scenario datasets for training SNNs.

IV. METHODS

In this section, we present a comprehensive methodology that enables the conversion of various datasets into temporal-encoded spikes suitable for SNN processing. The proposed approach combines techniques of data preprocessing, feature extraction, and network configuration to ensure the compatibility of the converted spikes with the SNN architecture.

A. Datasets

1) *MNIST*: MNIST is a standard database consisting of grayscale images of handwritten digits. It contains 60,000 training images and 10,000 test images, with each image having a size of 28×28 pixels. One of the advantages of using this database is that it requires minimal data preprocessing since most machine learning libraries have built-in support for MNIST. Although MNIST is not encoded in spikes, it can be utilized for SNN development in such a way that classical ANNs are first trained on MNIST and then converted to SNNs. It should be noted that MNIST does not include a separate validation set. If needed, a few samples from the training set have to be separated for such a purpose.

2) *N-MNIST*: Neuromorphic MNIST (N-MNIST) [4] is a spike-based representation of the MNIST database. It captures the dynamics of handwritten digits using a Dynamic Vision Sensor (DVS) camera. N-MNIST consists of the same number of samples as MNIST. Each sample is encoded as a binary file, storing pixel index (x, y), event type (ON or OFF), and the event timestamp. The events represent changes in light intensity. N-MNIST provides dynamic sequences with a duration of 300 ms and a resolution of 34×34 pixels. The motion in N-MNIST is inspired by the saccadic eye movement, featuring rapid movements in three directions lasting 100 ms each. This database enables the exploration of SNNs and their performance on tasks involving temporal information, serving as an alternative to static image-based databases.

B. Conversion Procedure: Transforming Data for Effective SNN Training

The first step for enabling SNN training on the MNIST dataset is to encode the data samples into time-distributed spike sequences.

The conversion pipeline consists of the following steps:

- Loading the image data
- Preprocessing the image data
- Scaling the image data
- Converting the image data into spike sequences
- Saving the converted spike sequences for later use in training

To implement this spike encoding procedure, several libraries written in Python have been used: *Pytorch* and *Pytorch-vision* for MNIST loading, *snnTorch* library for

temporal encoding and *h5py* library for saving the converted spike sequences.

The next step was to preprocess the image data to be compatible with the *snnTorch* library. One main preprocessing step was to scale the image values between 0 and 1. This step was mandatory as it was a requirement from the *snnTorch* library. One can do multiple preprocessing operations step such as resizing, rotating, etc, at this step.

The scaled values of the data are then used as an input for the spike generation module of the *snnTorch* library (this module enables to use several encoding schemes, either of rate or temporal types). The simulation time of the temporal encoded samples may be specified to a desired value.

The output from the spike generation module is an array of discrete values 0 and 1, in which value of 1 represents a spike. The array is a two-dimensional array of [T, U] size, in which the first dimension represents the simulation time index and the second one corresponds to the feature number, T is the total number of simulation time steps, and U represents the number of features (which matches SNN input layer size).

The final step is to save the generated spike data for the later use in SNN training. A common practice in neuromorphic datasets is to save the data in the form of arrays that correspond to: coordinates (x,y), spike times, and labels. Each event (spike) has a coordinate that corresponds to the index of the input neuron firing a specific spike, and a timestamp. In addition, a label is assigned to the whole sample. We have followed this common practice and reformulated the encoded spike data into corresponding arrays as mentioned above.

In our experiments, we have chosen a temporal encoding scheme, because of lower number of spikes needed to carry the information. This encoding scheme belongs to the group of temporal encoding schemes. As opposed to the rate encoding schemes, temporal encoding schemes contain a smaller number of spikes. In rate encoding schemes the average number of spikes represents the information. However, in temporal encoding schemes, the precise timing of a single spike carries the information. There are several temporal encoding schemes available. [25]. In this experiment, we applied the Time-to-First-Spike encoding procedure (T2FS) [26]. In the case of T2FS, the information is carried in the time of the first spike from the beginning of the simulation. It was experimentally proven, that tactile systems (e.g., at the fingertips) use a similar scheme to encode and transmit information about touch. Also, it has been suggested that the first spike carries twice as much information compared to rate encoding [27].

Lower number of spikes has a positive impact on overall hardware requirements, especially on the size of the batch in GPU. Also, this lower number of spikes reflects lower requirements for energy if a SNN processing this dataset would be implemented on hardware. The simulation time of samples was set to 30 ms as opposed to 300 ms simulation time of N-MNIST samples. The shorter sample time was selected to investigate the ability of SNNs to process samples with short sample time.

The encoding procedure of the *spikegen* module returns a multi-dimensional array with values 0 and 1. Note, due to

differences between input structure of the library used for SNN training with converted samples and the output structure of the *spikegen* module, the structures need to be reorganized to fit each other. The proposed pipeline block diagram is shown in Fig. 2.

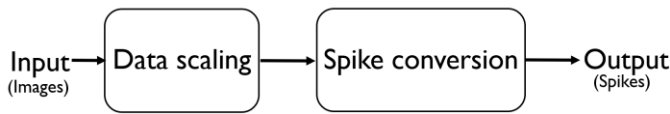


Fig. 2. Proposed pipeline block diagram.

C. Baseline System

The Spiking Convolutional Neural Network (SCNN), that is similar to the structure published in [13] was chosen as a baseline architecture. The architecture of the baseline system is shown in Table II.

The baseline model was trained on N-MNIST. The output layer of the model uses the rate encoding method, where the neuron with a higher spike rate is selected as the neuron representing the class. The neurons were trained to spike 60 times for the representing (true) class and 10 times otherwise (false class). The simulation time of SNN was 300 ms. The trained SNN was able to classify samples with processing delay of 150 ms. Although the process was more biologically plausible, its computational cost may be a disadvantage.

TABLE II. THE ARCHITECTURE OF THE BASELINE SYSTEM

Layer	Parameters
Input	34×34×2
Conv1	12 kernels (5×5)
Delay1	-
Pool1	2×2
Delay2	-
Conv2	64 kernels (5×5)
Delay3	-
Pool2	2×2
Delay4	-
Fc1	10

D. The Proposed System

As an alternative we propose an approach using a software conversion of the MNIST dataset that may be widely available and without the need of a specialized hardware (in contrast to N-MNIST).

The proposed architecture is again a SCNN similar to the baseline, except the input layer. The size of the input layer corresponds with the size and format of the MNIST image data. The structure of the model consists of convolutional layers followed by pooling. After each layer also a time delay

layer is applied. The delay layer is used as a special layer in SNNs. The SCNN output layer is a spiking fully connected layer with 10 neurons corresponding with the number of classes to be recognized. The whole architecture is shown in Table III.

The proposed system was trained from scratch for 80 epochs. The time of the whole training process was around 2.5 hours (using NVIDIA RTX 2060TI with 6GB of GPU memory). We used the Adam optimizer with starting learning rate of 0.001. The learning rate parameter was modified by the *ReduceLROnPlateau* learning rate scheduler, which modified the value of the learning rate based on criteria. The batch size of both train and test subsets was set to 32.

To implement and train the proposed SCNN on the converted MNIST dataset we applied the *Pytorch* library along with the *slayerPytorch* library that includes an implementation of SNN training using *Pytorch*. SNNs implemented using this library consists of Spike Response Model spiking neurons. The architecture is trained using the Spike SLAYER algorithm. This algorithm uses a surrogate gradient approach to overcome difficulties with training SNNs. This library contained all building blocks to create a SNN. This includes special layers that were made of spiking neurons, spike processing and the surrogate gradient method. There are more parameters that were needed to be configured. Most of these parameters were related to the *slayerPytorch* library and were used to control the simulation of SNN. We used similar parameters as the baseline, except that our samples had different simulation length. Note, our proposed model was trained for temporal encoding, where a lower number of spikes were needed for the model to classify a sample than in the case of the rate encoding that was used in the baseline.

TABLE III. ARCHITECTURE OF THE MODEL

Layer	Parameters
Input	28×28
Conv1	12 kernels (5×5)
Delay1	-
Pool1	2×2
Delay2	-
Conv2	64 kernels (5×5)
Delay3	-
Pool2	2×2
Delay4	-
Fc1	10

V. EXPERIMENTAL RESULTS AND DISCUSSION

The trained architecture was able to achieve accuracy of 98.79% on the MNIST test set. The changes in accuracy and loss during the training process are shown in Fig. 3 and Fig. 4.

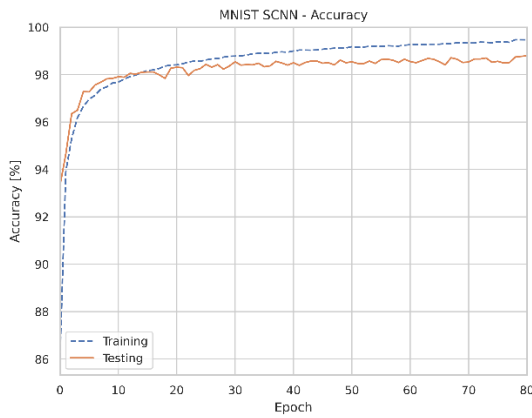


Fig. 3. Accuracy on subsets during the training process.

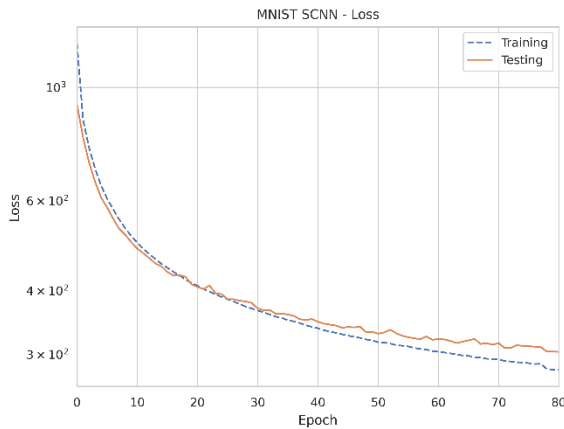


Fig. 4. Loss of the model during the training process.

The accuracy of the baseline system on the N-MNIST test set was 99.2%. The performance of the baseline system is comparable with other systems trained on N-MNIST and published. Although, the accuracy of our trained model was not higher than the baseline system, it is still a competitive number.

The improvement of our model is in spike efficiency. Our trained model was able to classify MNIST samples with only average of 27 spikes on the output layer during the simulation time. The baseline system for N-MNIST used 150 spikes on average during the simulation time. Our proposed system uses a more energy-efficient temporal encoding method, while the baseline system used rate encoding. Also, the number of spikes that are on the input of the SNN is lower. Baseline system for N-MNIST used around 4100 spikes on average, while our system used only 784 spikes. When encoding an image, the spikes are first occurring on the indexes of pixels with higher brightness intensity, while spikes on indexes of pixels with lower brightness intensity occur later. Note, each represented pixel contains only one spike. This means, that for MNIST images with a resolution of 28x28 pixels only 784 spikes occur for each image. Such a number of spikes are much smaller than in the case of the N-MNIST dataset, in which each sample around 4100 spikes on average.

Our proposed model is more energy efficient not only on the input spikes, but also on the output spikes. The comparison

of the baseline system and our proposed system is summarized in Table IV.

TABLE IV. COMPARISON OF THE BASELINE SYSTEM AND OUR PROPOSED SYSTEM

	Baseline system	Proposed system
Dataset	N-MNIST	MNIST (converted with the proposed pipeline)
Encoding	Rate encoding	Temporal encoding
Number of spikes on input (average)	4100	784
Number of spikes on output	150	27
Accuracy	99.2%	98.79%

Although our experiments are carried out on MNIST benchmark dataset, the methodology we present is versatile and applicable to other image datasets as well as to diverse data modalities, including audio or biological signals (e.g. in the form of spectrograms).

VI. CONCLUSION

In this paper we have presented a software conversion process that uses an energy efficient temporal encoding method to convert static image data into a format of spikes distributed in time. The proposed method was compared with a baseline method that used a specialized hardware for converting the same dataset. The baseline system used rate encoding. The functionality of the proposed method was examined on SNN that used a similar architecture to the baseline system. The results of the proposed solution in terms of accuracy were competitive with the baseline. However, the SNN trained using the proposed temporal encoding needs a significantly lower number of spikes in both input and output spike trains to correctly classify the dataset.

We envisage that the proposed pipeline will not only facilitate improved training and testing of SNNs but also inspire the development of larger datasets that cover a broader application domain. Through these efforts, we attempt to unlock the immense potential of SNNs and neuromorphic computing while advancing our understanding of brain-inspired computation.

The future improvements for the proposed pipeline may be in experiments with more datasets to be converted into spike trains and then used for training SNNs. The length of samples in time may play a role in the accuracy of the trained model and would need to be further investigated.

ACKNOWLEDGMENT

This work was supported by the Slovak Grant Agency KEGA under contract no. KEGA 008ZU-4/2021 and also by the ERDF project of Operational Programme Integrated Infrastructure, ITMS2014+ code 313011ASK8.

REFERENCES

- [1] S. Ghosh-Dastidar and H. Adeli, 'Spiking neural networks', *Int. J. Neural Syst.*, vol. 19, no. 04, pp. 295–308, Aug. 2009, doi: 10.1142/S0129065709002002.

- [2] S. Thorpe, A. Delorme, and R. Van Rullen, 'Spike-based strategies for rapid processing', *Neural Netw.*, vol. 14, no. 6, pp. 715–725, Jul. 2001, doi: 10.1016/S0893-6080(01)00083-1.
- [3] T. Zhang, M. R. Azghadi, C. Lammie, A. Amirsoleimani, and R. Genov, 'Spike sorting algorithms and their efficient hardware implementation: a comprehensive survey', *J. Neural Eng.*, vol. 20, no. 2, p. 021001, Apr. 2023, doi: 10.1088/1741-2552/acc7cc.
- [4] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, 'Converting Static Image Datasets to Spiking Neuromorphic Datasets Using Saccades', *Front. Neurosci.*, vol. 9, 2015, Accessed: Feb. 07, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2015.00437>
- [5] 'DVS128 Gesture Dataset - IBM Research'. <https://research.ibm.com/interactive/dvsgesture/> (accessed Feb. 07, 2023).
- [6] 'TIDIGIT Spikes Dataset', Google Docs. https://docs.google.com/document/d/1Uxe7GsKKXcy6SIDUX4hoJVAC0-UkH-8kr5UXp0Ndi1M/edit?usp=embed_facebook (accessed Feb. 07, 2023).
- [7] P. Lichtsteiner, C. Posch, and T. Delbruck, 'A 128times 128 120 dB 15 μ s Latency Asynchronous Temporal Contrast Vision Sensor', *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb. 2008, doi: 10.1109/JSSC.2007.914337.
- [8] J. Feng and David Brown, 'Integrate-and-fire Models with Nonlinear Leakage', *Bull. Math. Biol.*, vol. 62, no. 3, pp. 467–481, May 2000, doi: 10.1006/bulm.1999.0162.
- [9] W. He et al., 'Comparing SNNs and RNNs on Neuromorphic Vision Datasets: Similarities and Differences'. arXiv, May 02, 2020. Accessed: May 19, 2023. [Online]. Available: <http://arxiv.org/abs/2005.02183>
- [10] G. K. Cohen, G. Orchard, S.-H. Leng, J. Tapson, R. B. Benosman, and A. van Schaik, 'Skimming Digits: Neuromorphic Classification of Spike-Encoded Images', *Front. Neurosci.*, vol. 10, 2016, Accessed: May 19, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2016.00184>
- [11] Y. Wu, L. Deng, G. Li, J. Zhu, Y. Xie, and L. Shi, 'Direct Training for Spiking Neural Networks: Faster, Larger, Better', *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, Art. no. 01, Jul. 2019, doi: 10.1609/aaai.v33i01.33011311.
- [12] B. Ramesh, H. Yang, G. Orchard, N. A. Le Thi, S. Zhang, and C. Xiang, 'DART: Distribution Aware Retinal Transform for Event-Based Cameras', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2767–2780, Nov. 2020, doi: 10.1109/TPAMI.2019.2919301.
- [13] S. B. Shrestha and G. Orchard, 'SLAYER: Spike Layer Error Reassignment in Time', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2018. Accessed: Feb. 17, 2023. [Online]. Available: <https://proceedings.neurips.cc/paper/2018/hash/82f2b308c3b01637c607ce05f52a2fed-Abstract.html>
- [14] A. Sironi, M. Brambilla, N. Bourdis, X. Lagorce, and R. Benosman, 'HATS: Histograms of Averaged Time Surfaces for Robust Event-Based Object Classification', in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 1731–1740. doi: 10.1109/CVPR.2018.00186.
- [15] C. Lee, S. S. Sarwar, P. Panda, G. Srinivasan, and K. Roy, 'Enabling Spike-Based Backpropagation for Training Deep Neural Network Architectures', *Front. Neurosci.*, vol. 14, 2020, Accessed: Jul. 28, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2020.00119>
- [16] Y. Bi, A. Chadha, A. Abbas, E. Bourtsoulatzé, and Y. Andreopoulos, 'Graph-Based Object Classification for Neuromorphic Vision Sensing', in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 491–501. doi: 10.1109/ICCV.2019.00058.
- [17] Y. Jin, W. Zhang, and P. Li, 'Hybrid Macro/Micro Level Backpropagation for Training Deep Spiking Neural Networks', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2018. Accessed: Jul. 28, 2023. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/hash/3fb04953d95a94367bb133f862402bce-Abstract.html
- [18] A. Yousefzadeh, G. Orchard, T. Serrano-Gotarredona, and B. Linares-Barranco, 'Active Perception With Dynamic Vision Sensors. Minimum Saccades With Optimum Recognition', *IEEE Trans. Biomed. Circuits Syst.*, vol. 12, no. 4, pp. 927–939, Aug. 2018, doi: 10.1109/TBCAS.2018.2834428.
- [19] Y. Wu, L. Deng, G. Li, J. Zhu, and L. Shi, 'Spatio-Temporal Backpropagation for Training High-Performance Spiking Neural Networks', *Front. Neurosci.*, vol. 12, 2018, Accessed: May 29, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2018.00331>
- [20] J. H. Lee, T. Delbruck, and M. Pfeiffer, 'Training Deep Spiking Neural Networks Using Backpropagation', *Front. Neurosci.*, vol. 10, 2016, Accessed: Jul. 28, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2016.00508>
- [21] Q. Liu, H. Ruan, D. Xing, H. Tang, and G. Pan, 'Effective AER Object Classification Using Segmented Probability-Maximization Learning in Spiking Neural Networks', *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 02, Art. no. 02, Apr. 2020, doi: 10.1609/aaai.v34i02.5486.
- [22] J. Kaiser, H. Mostafa, and E. Neftci, 'Synaptic Plasticity Dynamics for Deep Continuous Local Learning (DECOLLE)', *Front. Neurosci.*, vol. 14, 2020, Accessed: Jul. 28, 2023. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fnins.2020.00424>
- [23] B. Cramer, Y. Stradmann, J. Schemmel, and F. Zenke, 'The Heidelberg Spiking Data Sets for the Systematic Evaluation of Spiking Neural Networks', *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 2744–2757, Jul. 2022, doi: 10.1109/TNNLS.2020.3044364.
- [24] A. Amir et al., 'A Low Power, Fully Event-Based Gesture Recognition System', in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 7388–7397. doi: 10.1109/CVPR.2017.781.
- [25] J. K. Eshraghian et al., 'Training Spiking Neural Networks Using Lessons From Deep Learning'. arXiv, May 15, 2023. Accessed: May 29, 2023. [Online]. Available: <http://arxiv.org/abs/2109.12894>
- [26] B. Rueckauer and S.-C. Liu, 'Conversion of analog to spiking neural networks using sparse temporal coding', in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018, pp. 1–5. doi: 10.1109/ISCAS.2018.8351295.
- [27] H. P. Saal, S. Vijayakumar, and R. S. Johansson, 'Information about Complex Fingertip Parameters in Individual Human Tactile Afferent Neurons', *J. Neurosci.*, vol. 29, no. 25, pp. 8022–8031, Jun. 2009, doi: 10.1523/JNEUROSCI.0665-09.2009.

A Secure and Scalable Behavioral Dynamics Authentication Model

Idowu Dauda Oladipo¹, Mathew Nicho², Joseph Bamidele Awotunde³,
Jemima Omotola Buari⁴, Muyideen Abdulraheem⁵, Tarek Gaber⁶

Department of Computer Science, University of Ilorin, Ilorin, Nigeria^{1, 3, 4, 5}
Research and Innovation Center, Rabdan Academy, Abu Dhabi²

School of Science, Engineering & Environment, University of Salford, Manchester, United Kingdom⁶

Abstract—Various authentication methods have been proposed to mitigate data breaches. However, the increasing frequency of data breaches and users' lack of awareness have exposed traditional methods, including single-factor password-based systems and, two-factor authentication systems, to vulnerabilities against attacks. While behavioral authentication holds promise in tackling these issues, it faces challenges concerning interoperability between operating systems, the security of behavioral data, accuracy enhancement, scalability, and cost. This research presents a scalable dynamic behavioral authentication model utilizing keystroke typing patterns. The model is constructed around five key components: human-computer interface devices, encryption of behavioral data, consideration of the authenticator's emotional state, incorporation of cross-platform features, and proposed implementation solutions. It addresses potential typing errors and employs data encryption for behavioral data, achieving a harmonious blend of usability and security by leveraging keyboard dynamics. This is accomplished through the implementation of a web-based authentication system that integrates Convolutional Neural Networks (CNNs) for advanced feature engineering. Keystroke typing patterns were gathered from participants and subsequently employed to evaluate the system's keystroke timing verification, login ID verification, and error handling capabilities. The web-based system uniquely identifies users by merging their username-password (UN-PW) credentials with their keyboard typing patterns, all while securely storing the keystroke data. Given the achievement of a 100% accuracy rate, the proposed Behavioral Dynamics Authentication Model (BDA) introduces future researchers to five scalable constructs. These constructs offer an optimal combination, tailored to the device and context, for maximizing effectiveness. This achievement underscores its potential applications in the realm of authentication.

Keywords—Behavioral authentication; keystroke dynamics; human-computer interface; two-factor authentication

I. INTRODUCTION

Stolen credentials remain the most frequent means by which hackers commit data breaches. In general, 80% of hacking-related breaches involve brute force or stolen credentials, and 37% of all breaches involve the use or theft of credentials [1]. Ensuring data protection and privacy through effective and efficient authentication methods is not only a priority for individuals, organizations, and countries but also a context for promoting incremental and radical innovation. Data-protection authorities must acquire new technologies and

use them effectively to regulate personal information practices so as to meet present and future challenges, for the technology advances in step with the security threats [2-3]. Hence, cybersecurity attacks have been increasing exponentially and rendering existing detection mechanisms insufficient [4]. Cyber-criminals exploit security flaws to access users' data and privileges [5], and, in turn, tactics such as authentication serve to restrict and control access to unauthorized users [6].

More than 555 million passwords have been obtained through data breaches and exposed in the public domain [7]. Some 27% of those surveyed in a Google poll admitted attempting to guess the passwords of others, 17% of whom claimed to have succeeded [8], and, according to one report, 80% of hacking incidents are enabled by stolen and reused login information, 81% of which at the company level are caused by the many poor passwords among the 300 billion in use [9]. Verizon's 2022 Data Breach Investigations Report identified passwords as a frequent weak link in cybersecurity, with 80% of all data breaches worldwide again being associated with passwords [10].

Multifactor authentication (MFA) can block more than 99.9% of account compromise attacks [11], and two-factor authentication (2FA) is currently among the primary mechanisms for defending against password attacks [12], especially those involving phishing and password reuse [13]. Biometric authentication (BA) based on keystrokes is considered more reliable than these traditional means of authentication because of its novelty and low intrusiveness [14]. A National Bureau of Standards study found keystroke dynamics (KD) to be a reliable and accurate BA method, achieving at least 98% accuracy [15].

The overwhelming focus in the scholarship for several decades has been on the legal and technological dimensions of the challenges associated with data protection [2]. Among the world's privacy and security laws, the European Union General Data Protection Regulation (GDPR), adopted in 2016 and enacted on May 25, 2018, is the strongest [16], and a crucial aspect of compliance with this regulation is adequate authentication and authorization [17]. Accordingly, biometrics generated through KD has become a viable option to authenticate users for both security and surveillance purposes [18] given the minimal implementation cost of KD and lack of a need for special hardware since the gathering of typing data

is reasonably straightforward, requiring no additional effort by the user [19].

Accordingly, we propose a scalable dynamic behavioral authentication model incorporating future trends in information systems, specifically, trends in or toward (1) devices, (2) dynamic encryption standards, (3) assessing the emotional state of the authenticator, (4) the ubiquitous access to and usage of devices, and the option to (5) incorporate emerging technologies and solutions for low cost and increased functionality. The result, validated based on these five constructs, is a cost-effective and non-intrusive MFA system method to augment users' authentication and ensure data security while efficiently handling their typing errors [19]. The use of keystrokes itself validates the model owing to its simplicity and its ability to integrate seamlessly with passwords. Simply put, KD is an extremely useful method for BA because it is extremely difficult to impersonate [20].

The main contributions of this paper are:

- a dynamic behavioral authentication model incorporating five scalable constructs namely HCI devices, encryption, user profiles, ubiquitous, and the options for cost effective applications,
- scalability in all five constructs of the model that offers multiple avenues for future research on combinations of the components of the constructs,
- a secure and efficient authentication system based on users' unique key-typing patterns with an error-correction feature,
- a cost-effective way to implement BA compared with other methods (which may require a combination of hardware and software),
- alignment of the proposed system with the relevant data protection and privacy regulations (e.g., the GDPR) to assist organizations with compliance.

In the remainder of the paper, Section II presents our justification for using 2FA and MFA and our analysis of attacks followed by a review of the research on the application of behavioral authentication in KD. Section III presents the methodology and Section IV discusses the application of KD and the resulting model. We discuss our conclusions in Section V.

II. LITERATURE REVIEW

The following discussion surveys current issues relating to password authentication, the use and usability of multifactor authentication, the ease of password attacks, and behavioral authentication methods.

A. Challenges in Password Authentication

Password-based authentication has been fraught with multiple issues. Firstly, the resilience of diverse operating environments is considered a significant technical challenge for future biometric systems. [21]. Secondly, the endeavor to develop efficient and secure biometric authentication systems that can withstand impersonation attacks, guarantee the non-reversibility of biometric templates, and safeguard the privacy

of personal information is critical [22]. Thirdly, the challenge lies in establishing appropriate policies and laws to prevent the indiscriminate use of biometric data [23]. From a keystroke recognition perspective, there is a necessity for the development of technology to enhance accuracy [24]. Lastly, a majority of the current research on keystroke dynamics revolves around free text (emphasizing n-graphs). This not only poses challenges in improving accuracy but also frequently requires a substantial amount of time to construct user models [25]. The suggested research surmounts these constraints by achieving a 100% accuracy enhancement and employing the Homomorphic Public Key Encryption technique for training and predicting encrypted data. This approach guarantees privacy while expediting training without sacrificing usability.

However, this can be enhanced through multifactor authentication systems that combine passwords, biometrics, and OTP verification [26]. The study [27] observed that, to control access to data, most systems rely on usernames and passwords for authentication, which are both convenient and insecure because they can be quickly entered into an online application or service [28]. Since the human capacity for information processing is limited, users face difficulties in remembering and matching their passwords and, therefore, often use either easy-to-guess passwords or complex passwords that are hard to remember [29]. Many internet applications, such as remote logins, for government organizations, private corporations, database management systems, and school systems are based on password authentication, but the current internet environment is vulnerable to replay, guessing, modification, stolen verifier, and other types of attacks [30]. Regardless of the complexity of a password string, passwords are considered a weak form of authentication because they can be shared, stolen, forgotten, or hacked [31]. Despite several major problems that have been identified with alphanumeric passwords, they are still used to protect both low-sensitivity and highly sensitive information [29] and remain the most widely used method of end-user authentication [32]. Since password-based user authentication methods provide only partial protection from hackers and intruders, additional authentication should be applied [33].

B. Multifactor Authentication (MFA)

MFA has emerged as a substitute mechanism to increase security by demanding that users provide more than one factor of authentication (i.e., in addition to a password) [32]. This layered approach to authentication provides more robust protections for users and minimizes the risks of breaching security [5] because a hacker must penetrate multiple layers of security to gain access to an MFA-enabled system. The various types of MFA present various security issues but combining them with password-based authentication systems can greatly improve the credibility of a user's login and complicate access for intruders [34].

Factors such as "what the user knows" (inherent), "what the user has" (possession), and "what the user is" (biometric) have served as additional authentication methods [32]. In this regard, biometric features unique to individuals, ranging from physiological characteristics (e.g., fingerprints and iris, hand, and face patterns) to behavioral characteristics (e.g., KD,

mouse movements, gait, and handwriting), have served to identify users [3]. Many such biometric approaches tend either to be expensive or to place heavy demands on computer hardware, making them inappropriate for most users [35].

C. Review of Authentication Methods

Human interface devices such as keyboards and mouse have been used extensively to help users interact with computer devices to increase their productivity. In this respect, the analysis of computer users' typing behavior (i.e., KD) is described as a behavioral biometric (BB) that can be analyzed and measured to improve cyber defenses [36]. The analysis of KD through BB (Fig. 1) focus on human-computer interaction (HCI) involving mouse movements and keyboard strokes [37]. The method proposed concentrates on a subcategory of HCI, wherein the innovative secure authentication approach relies on users' behavioral patterns manifested through keystrokes.

While extensive and valuable research on KD has facilitated authentication systems, challenges relating to the current user-authentication features include: -

- the increased use and introduction of HCI devices,
- the need to secure a behavioral and biometric database through relevant encryption standards,
- the ubiquitous and universal nature of devices that require cross-platform functionality,
- users' input errors during UN and PW entry linked to emotional states and secure web forms that restrict genuine access after a limited set of related errors,
- the availability of machine learning- and web-based solutions and hardware for authentication mechanisms, and
- biometric systems that can be forced on victims (unconsciously or under duress).

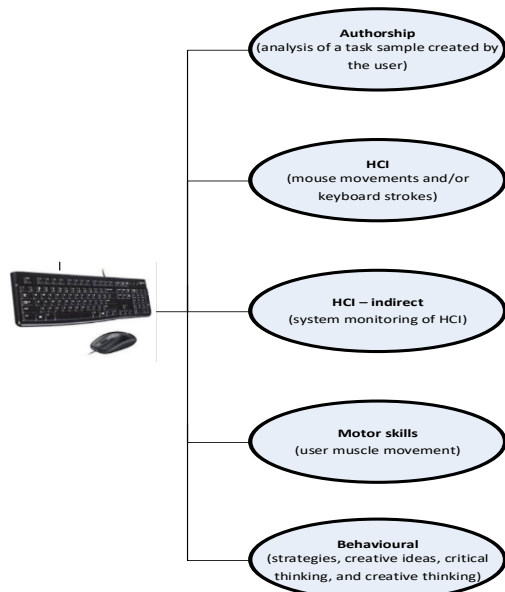


Fig. 1. Classification of behavioral biometrics [adapted from 37].

A lightweight, web-based secure authentication system that can be seamlessly integrated into multiple browsers based on behavioral KD meets these challenges. Research on KD using the keyboard and/or mouse in the literature includes authentication studies using ML, a combination of machine learning (ML) and web-based solutions, and hardware tokens. A Google Scholar title search using “keyboard dynamics” and “user authentication” generated seven relevant results, and a further search for these phrases anywhere in articles published since 2016 generated 62 results. These search results indicate that academic researchers exploring the application of KD in systems and device authentication (Table I) have been focusing on:

- the use of ML techniques for enrollment and verification with KD from live users and publicly available datasets based on keyboard and/or mouse dynamics with a near-zero error rate,
- proposing and experimenting with web-based applications that incorporate KD, mouse dynamics, a combination thereof, a combination of KD and biometrics, and behavioral features to achieve an accuracy of 97.96%,
- methodologies and experimental applications using KD and a combination of KD and mouse dynamics with hardware and software to achieve an accuracy of 97%, and
- Client-server KD systems requiring both hardware and software, along with peripherals like sensors, cameras, and hardware tokens.

D. Behavioral Authentication using ML

Several studies have described the successful or near-successful use of keystroke and/or mouse dynamics for authentication with public data sets as well as live users, mostly from a back-end server perspective. The author in [38] created a system using an ML classifier beginning with a support vector machine (SVM) room for each user that accesses data from 34 user environments and was able to accept real users with an extremely low error rate (nearly 0%). These findings demonstrate that a one-class SVM may serve as a tool for the continuous user analysis and validation of button dynamics through the formulation of plans and the choice of the proper kernel parameter. With histogram gradient boosting as the primary classifier for the training and testing phase on a keystroke benchmark dataset, [19] achieved 97.96% average accuracy and an average equal error rate (EER) of 0.014 across all subjects, thus outperforming all previous advances in both ML and deep learning approaches. The researcher in [39] used the confidence interval and k-means clustering to demonstrate the success of trajectory dissimilarity, achieving a higher accuracy rate, 96%, than other techniques. While trajectory dissimilarity uses KD for the usernames confidence interval and k-means clustering uses KD for passwords, another significant finding from the study is the use of the former to protect users from account lockout attacks (i.e., exploiting the lock-out of accounts after a defined number of incorrect password attempts).

TABLE I. REVIEW OF EXPERIMENTAL RESEARCH USING KEYBOARD AND / OR MOUSE DYNAMICS

	Author	ML	Live users	Public dataset	Error-correction method	Web-based	Accuracy (%)	Error rate
	Ibrahim et al., 2023	Yes	No	Yes	No	No	97.96	X
	Anusas-amornkul & Wangsuk, 2015	Yes	No	Yes	No	No	96.00	X
	Quimatio, Njike, & Nkenlifack, 2022	Yes	No	Yes	No	No	95.65	X
	Raul, Shankarmani, & Joshi, 2020	Yes	Yes	No	No	No	90.50	X
	Kar, Bamotra, Duvvuri, & Mohanan, 2023	Yes	Yes	No	No	No	95.05	X
	Phadol, 2022	Yes	No	Yes	No	No	X	Near 0
Machine Learning focus	Shi, Wang, Zheng, & Cao, 2022	Yes	Yes	No	No	No	89.22	FAR of 11.27%; FRR of 10.25%
	X. Wang et al., 2022	Yes	Yes	No	No	No	84.00	X
	Jadhav et al., 2017	Yes	Yes	No	No	No	X	FAR of 1%; FRR 4%
	Stragapede et al., 2022	Yes	No	Yes	No	No	X	EER 03.25%
	Piugie, Di Manno, Rosenberger, & Charrier, 2022	Yes	No	Yes	No	No	X	EER of 04.49%
	Gupta et al., 2015	Yes	Yes	No	No	?	X	FAR of 05.4%; FRR of 09.2%
	Alshanketi et al., 2016	Yes	Yes	Yes	Yes	No	X	X
	Bhattacharya et al., 2022	Yes	Yes	No	No	Yes	87.00	X
	Yang et al., 2023	Yes	No	Yes	No	No	X	X
	(Shekhawat & Bhatt, 2022)	Yes	No	Yes	No	No	97.00	X
Web-based focus	Zaidan et al., 2017	No	Yes	No	No	Yes	Positive results	EER of 2.3%
	Boakye Osei, Opanin Gyamfi, & Okoe Alhassan, 2020	No	Yes	No	No	Yes	X	FER of 4%; FRR of 6%; FAR of 1%
	Kang & Kim, 2023	No	Yes	No	No	Yes	X	Low FAR and FRR values.
	Siti Rahayu et al., 2020	No	Yes	No	No	Yes	X	Low ERR
	Rahman, Neupane, Zaiter, & Hossain, 2019	No	Yes	Yes	No	Yes	X	EER of 10.50%
	Vasyl, Sharapova, Ivanova, Denis, & Yuliia, 2017	No	Yes	No	No	Yes	X	X
	Cockell & Halak, 2020	**	Yes	No	No	No	X	Error rate of 4.5%
	Proposed Study	YES	Yes	No	Yes	Yes	100	X

*Natural language processing; **Portable hardware token. EER = equal error rate; FAR = false accept rate; FRR = false reject rate; FER = failure to enroll rate

Quimatio, Njike, and Nkenlifack [40] proposed an authentication method based on three bagging ensembles formed by an SVM, K-nearest neighbor (KNN), and decision tree classifiers, the outputs of which were merged using the CMU dataset to achieve 95.65% accuracy. Kar, Bamotra, Duvvuri, and Monahan [41] proposed KD for authentication by taking a dataset of 51 users who typed a password in eight sessions on alternate days to record fluctuations in their moods and implemented anomaly-detection algorithms based on distance metrics and ML algorithms, such as artificial neural

networks (ANNs) and convolutional neural networks (CNNs), to classify the users with 95.05% accuracy with an ANN with Negative Class.

Other studies of these issues include that of Raul, Shankarmani, and Joshi [42], who proposed combining non-conventional features with the conventional time-based features for user identification in static KD using ML classifiers and observed improvements in the false reject rate (FRR), false accept rate (FAR), and EER; their five ML algorithms determined that the logistic regression method

achieved 90.50% accuracy. Shi, X. Wang, Zheng, and Cao [43] proposed a user authentication method based on KD and mouse dynamics involving comparison of all of the representative time windows and dimensionality-reduction targets of the KD features to determine the parameters for ensuring the robustness of the algorithm and, using real-world setting, the HCI dataset achieved 89.22% accuracy in authenticating users, thus demonstrating the effectiveness of the algorithm. X. Wang, Shi, Zheng, Zhang, Hong, and Cao [44] presented a user authentication method that relies on scene-related and user-related features for user identification: first, features are extracted based on keystroke and mouse movement data; next, scene-related features are obtained that have a low correlation with scenes; lastly, scene-related and user-related features are fused to ensure their integrity. This proposed method has the advantage of improving user authentication accuracy in hybrid scenes, with an accuracy of 84%. Alshanketi, Traore, and Ahmed [31] presented an algorithm for handling typing errors in mobile keystroke BA combining timing- and pressure-based features. They used the random forest algorithm to classify and differentiate between trusted users and impostors based on a profile built for each user.

Researchers have also used ML to measure the error rates as success factors in authentication. Piugie, Di Manno, Rosenberger, and Charrier [45] proposed an approach based on the transformation of behavioral biometrics data (i.e., time series) into a 3D image that retains all of the characteristics of the behavioral signal and assists in training images based on CNNs and evaluates the performance of the system in terms of the EER based on a significant dataset, and they demonstrated the efficiency of the proposed approach on a multi-instance system. Mao, Wang, and Ji [46] combined keystroke content with keystroke time as the feature vector using a CNN to process the feature vectors and then input the normalized vector into the bi-LSTM network for training; they then tested this approach on an open data set and achieved an FRR, FAR, and EER of 3.09%, 3.03%, and 4.23%, respectively. Stragapede, Delgado-Santos, Tolosana, Vega-Rodriguez, Guest, and Morales [47] took into account the emotional and physical state of the authenticator and proposed a novel transformer architecture to model free-text KD performed on mobile devices using a publicly available Aalto mobile keystroke database, and they achieved experimental results that outperformed the current state-of-the-art systems, with an EER of 3.25% from only five enrollment sessions of 50 keystrokes each.

Jadhav Kulkarni, Shelar, Shinde, and Dharwadkar [3] proposed an ML-based authentication model that uses the static approach of keystroke dynamics to recognize and authenticate users accessing the system based on their unique keystroke profiles with respect to the flight, dwell, press, press-to-press, and release-to-release time and achieved an FAR and an FRR of 1% and 4%, respectively. Gupta, Khanna, Jagetia, Sharma, Alekh, and Choudhary [35] proposed a high-efficiency authentication system combining two methods to make keystroke biometrics less susceptible to forgery and more usable and reported that the system efficiently implemented secure authentication with the advantage of ease of

implementation since all that is required is the installation of software on any workstation. Yang et al. [55] focused on the text entered by the user and proposed contents and keystroke dual attention networks (ML) with pre-trained models for continuous authentication to address user-inputted “text” during keystrokes as an important asset beyond traditional KD characteristics, and their model achieved state-of-the-art performance on two datasets.

E. Web-based Behavioral Authentication

The research on various aspects of authentication systems has included web-based authentication, web-based keystroke authentication, portable tokens, and parametric approaches. Beginning with the first of these, to date, researchers have looked at web-based authentication. Thus, Bhattacharya, Trivedi, Obaidat, Patel, Tawar, and Hsiao [48] constructed a 2FA scheme for web users based on real-time KD by employing the KNN classification algorithm and achieved 87% accuracy over 146 testing samples and a recall value of 0.95, thus addressing the false-negative issue. Kang and Kim [49] used mouse dynamics and KD to identify personalized repeated user interface (UI) sequences with an Apriori algorithm based on the keystroke-level model of the HCI domain and validated the effectiveness of the system in complementing normal authentication through access testing with commercial applications that require intensive UI interactions.

Siti Rahayu, Guan, and Yusof [33] proposed an authentication system using KD in three stages—enrollment, verification/retraining, and client/server connection—and achieved a low error rate with just five users. Rahman [50] introduced a novel method utilizing KD as an additional validation layer in web-based applications such that users were prompted to type five words after registering their username and password (UN-PW), with the extracted features stored as a JSON object in the database. Vasyil, Sharapova, Ivanova, Denis, and Yulia [51] developed a web-based authentication system based on users’ keystroke features and suggested merging KD with other human features to achieve greater precision in authenticating users. Zaidan, Salem, Swidan, and Saifan [27] developed a web-based application for use in the study of the factors affecting KD in mobile systems that extracts and stores features such as the characters typed, key-hold latency, up-down latency, down-down latency, and overall latency; they then tested factors such as the device used for typing, the knowledge of the text, the mood of the user, and the complexity of the password on this dataset and achieved positive results. Boakye Osei, Opanin Gyamfi, and Okoe Alhassan [52] proposed a web-based keystroke login system using features such as dwell, flight, and locate to minimize error rates. Though the system achieved lower error rates than previous systems, it was limited to web-based formats and QWERTY keyboards. Aliksieiev, Strelitskiy, Gavva, Gorelov, and Synytsia [53] used a web-based application to gather and analyze users’ keystroke information based on a calculation of digraph timings and employed a non-parametric test to compare multiple datasets for situations in which distribution is difficult to determine and the sample is small.

From a cloud environment perspective, [54] used a combination of static authentication, click color-based dynamic authentication, and behavioral biometrics (keystroke with cryptographic encryption and a hashing technique) to achieve 96% accuracy and a decrease in false positives with an error rate of 3%. In [18] author developed a unique way to perform KD-based authentication with a keyboard and an array of pressure sensors that serves to develop unique user profiles that improve the suggested system's efficiency and, using a real-world dataset, achieved a 97% success rate in experiments. Using a natural language processing method, [56] introduced a portable hardware token for MFA using keystrokes to enhance authentication, but the proposed algorithm, though simple, achieved relatively low accuracy, but with a relatively high error rate. So, they suggested applying ML and considering close keystrokes to reduce authentication errors. ML techniques, especially CNN, have been effectively employed in behavioral authentication using gait recognition to extract high-level features from the input data [57]. Likewise, the random forest classifier and support vector machine have been utilized in behavioral authentication involving touch behavior (such as finger pressure, size, and pressure time) while tapping keys on smartphones. This approach achieved an accuracy of 97.80% employing just 25 features [58].

This review of the literature on KD reveals two major aspects of authentication, namely the use of multiple devices (KD and/or mouse dynamics, including touch screens) and the extensive use of ML and web-based solutions in authentication. However, there is a lack of emphasis on the applicability of Behavioral Authentication (BA) across different devices, the potential for scalability in encryption standards, the influence of various emotional states of users on behavioral data errors, cross-platform compatibility, and the need to incorporate upcoming applications and solutions for cost efficiencies. In this regard, a validated secure and scalable model can offer multiple alternatives for the five constructs.

The experiment phase of research centers on keystroke dynamics as a behavioral biometric, propelled by both the user's frequent utilization of this parameter and their proximity to the computing device. While biometric parameters can encompass both physiological and behavioral traits, behavioral aspects, such as the user's gait, interaction with the graphical user interface, haptic responses, programming style, registry access, system call logs, and mouse dynamics, offer advantages like persistent security, post-login authentication, ease of behavioral data collection, and the absence of a requirement for specialized hardware [37]. The evaluation metrics, namely FAR, FRR, ERR, and FER, are all zero due to the 100% acceptance rate. As a result, these evaluation metrics are not elaborated upon in the subsequent section. Additionally, the experimental results establish a foundation for the forthcoming model, which can be employed for future research endeavors. In terms of the number of users, researchers have extensively used ML classifiers on publicly available databases for error reduction and conducted experiments using live keystrokes with numbers of users ranging from five to hundreds in which the data revealed no change in effectiveness, even with few users.

III. METHODOLOGY

The discussion of the research methodology here includes the proposed framework (Fig. 2) and the methods employed during the development of the proposed system.

A. KD Authentication Framework

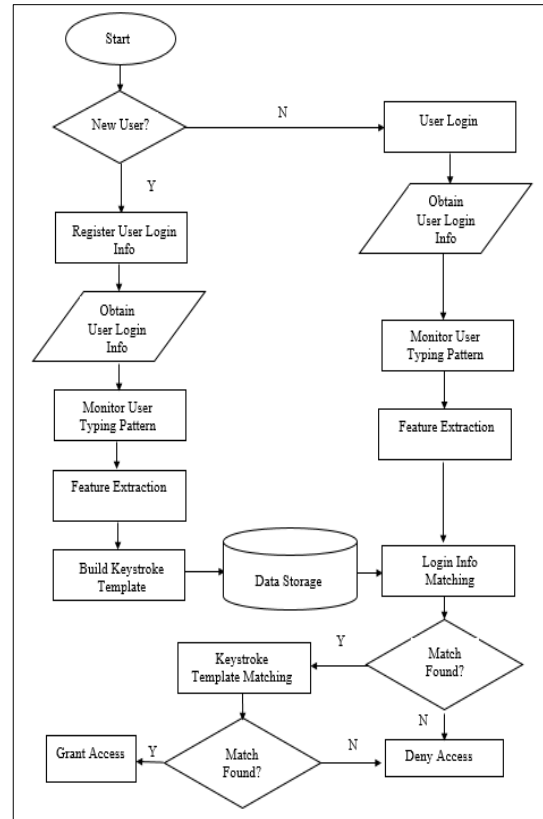


Fig. 2. Proposed framework.

B. Methods Employed

The method is a static approach that involves checking users' data during authentication. The experimental process starts with the training and testing phase and concludes with measures to handle errors. This section describes the environment under which our system was developed and evaluated.

1) *Experiment environment:* The system was developed using Python Programming 3.12 prerelease (2028-10 PEP 693) windows version running on a Laptop with specification (Core i5 with Intel Processor and 16 GB of RAM with Window 10 Operating version (Box 1).

2) *Dataset description:* The data were collected from five users. Each one was asked to type in the word "P@ssw0rd" in 10 trials. From each trail, four features (key press, key release, the characters typed by the user, and the total number of keys typed by the user) were extracted. Then three statistical features (average, mean and standard division) were added to each user's features. So, the total dataset used consists of 800 features (5x10x4x3).

3) *Classifier description:* In the classification phase, the CNN algorithm was used as follows. The features, extracted above, have been added at the fully connected layer of the CNN model. This further improved the classification accuracy.

Box 1 Python parameters

Python-CNN parameters tuning	Learning rate =
0.01, 0.001, 0.0001	
Neuron count= 8, 16, 12,	
Layer depth = 1, 2, 3,	
Kernel size = 8, 16, 12,	
Loss function = L2 loss, Binary cross-entropy,	
Epoch = 20, 50, 100	
Number of Hidden layers =2 Output layer=1	
Total number of layers =3	

4) *Security of the data:* Since the behavioral keystroke data from the five users is stored on a server, users are unable to modify the application due to encryption. The system utilizes keystrokes to ascertain whether a user is a legitimate individual or an imposter, both during the session and even after logging into the program. The Paillier cryptosystem algorithm, that comes under the category of the Homomorphic Public Key Encryption technique was utilized in the keystroke analysis training to predict encrypted data, ensure both privacy and accelerate training without compromising usability. It assumes a public key encryption technique that supports the following homomorphic property, as demonstrated by the equation: $E(m_1).E(m_2)$. This property can be stated as $E(m)^k = E(k.m)$ for various identical ciphertexts. For notation, we use the letter E to designate an encrypted value. For example, E(x) stands for the encryption of x, and (C) represents the encryption of the distance indication C.

a) *Training phase:* The training phase incorporates the enrollment and registration of new users, who are prompted to a sign-up page, thus enabling the system to acquire and store their data. This phase includes the following processes:

- Registering and Obtaining Users' Login Information: Every unique user signs up with a preferred UN-PW (with a minimum length) that is stored in the database.
- Monitoring Users' Typing Patterns: After registration, users are required to type a sequence of characters three times at various speeds while the system monitors their typing rhythm.
- Feature Extraction: Once users register, specific features are extracted while they are typing, including key press (how long a key is held down), key release (how quickly a pressed key is released), the characters typed, and the total number of keys typed.
- Building a Keystroke Template: After extraction of the behavioral features, the data are structured in an array to store the extracted features of each character independently of the other characters. To build a keystroke template, the fastest keystroke data typed serve as the upper bound and the slowest as the lower bound.

- Data Storage: After these features have been extracted and the template has been built, they are stored in the database along with users' UN-PWs for deployment to authenticate them when they try to log into the system.

b) *Testing phase:* The testing phase occurs during the login of existing users using the credentials registered in the enrollment process through the sign-up interface. This phase includes the following processes:

- Obtaining Users' Login Information: Each user must type in the previously chosen UN-PW, and the data are collected mainly for matching with the stored behavioral characteristics.
- Monitoring Users' Typing Patterns and Feature Extraction: As in the training phase, users are prompted to type a sequence of characters twice so that the system can extract matching keystroke data.
- Login Information and Keystroke Template-matching: The system matches the UN-PW to the pre-registered data stored in the database. If the data do not match, the user is given three more tries and then, if still unsuccessful, locked out. Authentication is ensured only after matching of the username, password, and behavioral keystroke data. If the average keystroke provided does not fall within the bounds described in 3.2.1, the user is prompted to try again in like manner as in the case of the UN-PW.

c) *Error handling:* Users may make mistakes when entering their passwords, thereby compromising the accuracy of the authentication system. To prevent this outcome, the keystroke data for each letter are gathered and stored separately based on how the letter is typed. This information is then organized into an array and stored in the database. If a user attempts to correct a mistake by deleting a character using the backspace key, the keystroke datum for the last letter entered is removed from the array.

IV. IMPLEMENTATION, RESULTS AND MODEL

A. Implementation

A web-based application served to implement the system, the main components of which are (1) a server-side programming language (NodeJs) for back-end functions, (2) a front-end language (HTML and CSS) for user interface, (3) a database (MongoDB) for storing users' profiles and keystroke data, (4) JavaScript for extracting keystroke data, and (5) a web browser. As the flow chart (Fig. 2) shows, the proposed system consists of two phases, registration (training) and login (testing). The following discussion describes each of these phases in turn.

1) *User registration:* As discussed, access to the system requires a login ID for every user (Fig. 3). The login data are collected and stored during the registration process. In this phase, data such as first and last name, username, email, and password are collected, as the figure shows, and then stored in the database.

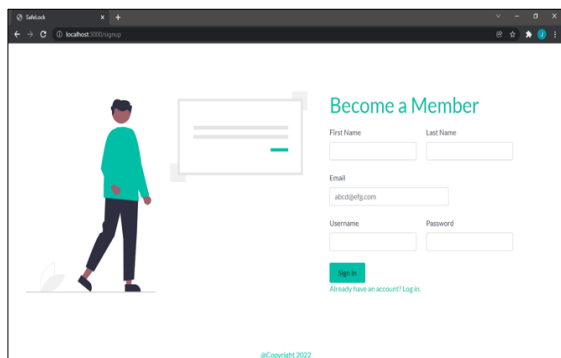


Fig. 3. User registration page.

2) *Generation of the keystroke template*: Once the login ID has been documented and stored in the database, the user is prompted through a screen as shown (Fig. 4).

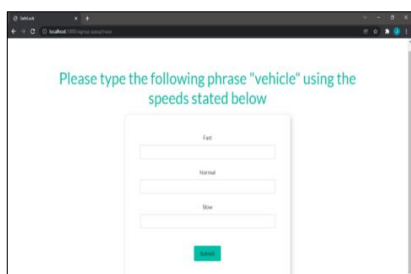


Fig. 4. Keystroke collection page.

Users are asked to type a randomly generated word in order to build a keystroke template. The features described above, such as the press and release of keys, which letters are typed, and the total number of characters typed, are extracted while the user types the word and stored independently in separate arrays. The keystroke data for the first letter occupy the first index of the array, the data for the second letter occupy the second index, and so on, as shown (Fig. 5).

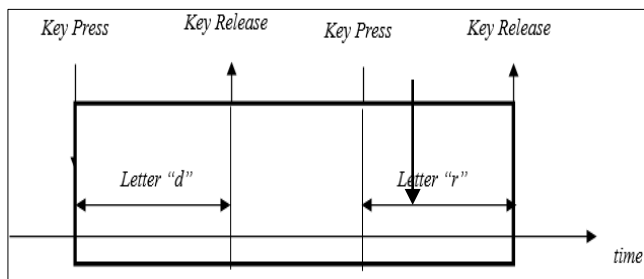


Fig. 5. Keystroke parameters.

The keystroke time KT , also referred to as the dwell time, is based on the user's typing rhythm for each letter of the randomly generated word and computed as

$$KT = K_{release} - K_{press}$$

where K_{press} is how long the user presses a key, $K_{release}$ is the time the user releases the pressed key, and $KT_{average}$ is the keystroke time of the phrase typed computed as the summation of the difference between the key press and key release times divided by the overall number of characters typed:

$$KT_{average} = \frac{\sum(K_{release} - K_{press})}{total\ number\ of\ letters\ typed}$$

Users are required to type the randomly generated word at various speeds as shown (Fig. 5) so that the system can collect their keystroke data at various speeds. The keystroke data at these speeds provide the bounds for each user's keystroke values, with the fastest keystroke serving as the lower bound and the slowest as the upper bound.

3) *Data storage*: Users' login ID and keystroke data are collected and stored in the database as shown (Fig. 6).

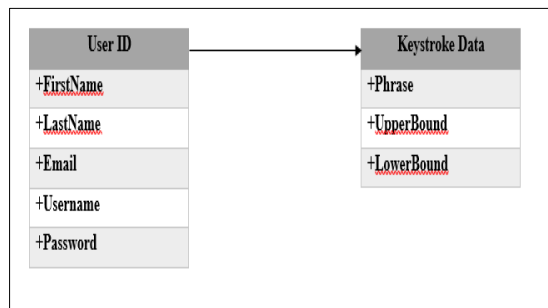


Fig. 6. Database model.

After users enter their login information on the registration page, the data are saved in the database, including the upper bound, the cumulative keystroke data, and the phrase used to log in. To minimize security risks, sensitive data, such as users' passwords and keystroke timing, are encrypted before storage in the database.

4) *User verification*: This process consists of the verification of users' login ID and keystroke timing as shown (Fig. 7).

a) *Login ID verification*: Once users try to access the system after registration, they are required to type in the credentials that they previously supplied, including the UN-PW, as shown in Fig. 7. Once the mandatory login data have been provided, the system matches the username and encrypted password by decrypting the latter to cross-check with the stored password hash for successful authentication, after which users are directed to the keystroke verification page.

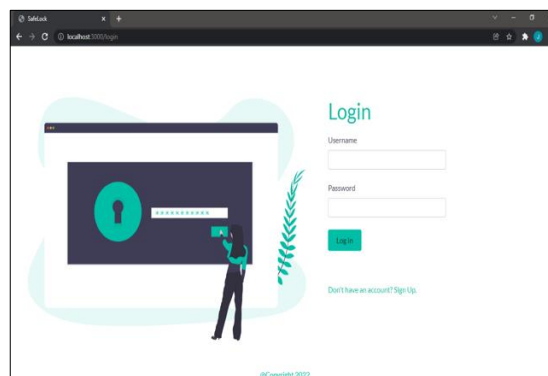


Fig. 7. Login ID verification.

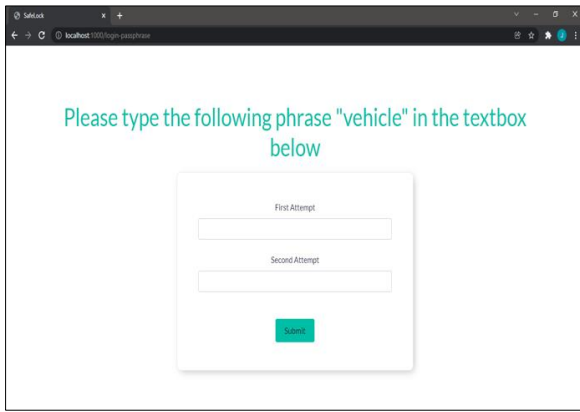


Fig. 8. Keystroke verification page.

b) *Keystroke timing verification:* On this form (Fig. 8), the user is required to enter the given phrase twice. During this process, the system computes the keystroke data for each letter typed and, once the user completes typing the first word or phrase, the average keystroke data. The process is then repeated with a second word or phrase. Once the user submits these data, the average of the keystroke data of the two words or phrases is computed and compared with the upper and lower keystroke bounds stored in the database, and the user is given access only if the average falls between them.

5) *Error handling:* Users are bound to make mistakes when typing. In keystroke-based systems, these mistakes can affect the accuracy of the system, but they can be managed efficiently. For example, in a situation in which a user types a sequence of characters, accidentally types a wrong character, and attempts to correct the error by tapping the backspace key. However, this attempt does not solve the problem of the mistyped character because the backspace key can be considered a character as well and the keystroke timing can be computed together with other characters. The proposed system incorporates an error-handling feature that deletes the last keystroke data accumulated after the backspace key is typed (Fig. 9).

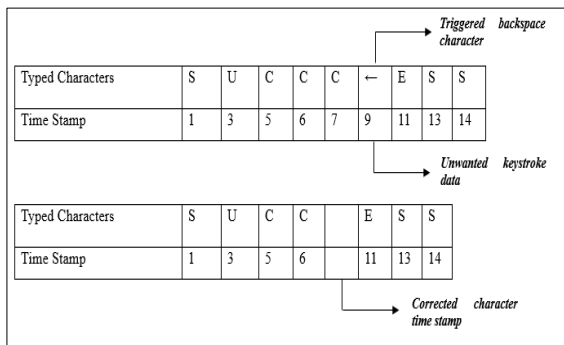


Fig. 9. Error handling using backspace.

B. Results and Discussion

To enroll in the system, each user types three words or phrases at fast, normal, and slow speeds. Taking into account the emotional state of authenticators, a rapid typing speed

indicates that they are in a hurry or excited, while a slow speed indicates feelings of tiredness or illness. This approach, then, establishes ranges of possible values for users' keystroke timing in order to authenticate them regardless of their emotional state. During the typing process, event listeners (key down and key up) served to capture the key press and the key release times. The system is designed so that, when users make mistakes and press the backspace character, instead of adding the last character typed, it deletes the last character from the array as shown (Fig. 10). This feature renders our system privacy-protected, as personal behavioral data is deleted when no longer required.

In the experiment, the system grants access to all users who have connected to it provided that their typing times fall within their ranges. Conversely, if the keystroke time is not within their ranges, users are not granted access. However, the system can also grant access to another entity if the stored behavioral values fall within the range (using the same UN-PW), and this feature can be enhanced by increasing the number of words/phrases (Fig. 8) entered into the systems as shown in Table II. Based on these features of our model, it can be noted that there is a time limit imposed on the storage of personal data, fulfilling one of the privacy protection requirements.

▶ (2) ['v', 'e']	signup-passphrase:95
▶ (2) [0.009, 0.029]	signup-passphrase:103
average:0.019	signup-passphrase:108
▶ (3) ['v', 'e', 'h']	signup-passphrase:95
▶ (3) [0.009, 0.029, 0.017]	signup-passphrase:103
average:0.018333333333333333	signup-passphrase:108
▶ (4) ['v', 'e', 'h', 'e']	signup-passphrase:95
▶ (4) [0.009, 0.029, 0.017, 0.021]	signup-passphrase:103
average:0.019	signup-passphrase:108
▶ (3) ['v', 'e', 'h']	signup-passphrase:86
▶ (3) [0.009, 0.029, 0.017]	signup-passphrase:87
▶ (4) ['v', 'e', 'h', 'i']	signup-passphrase:95
▶ (4) [0.009, 0.029, 0.017, 0.016]	signup-passphrase:103
average:0.017750000000000002	signup-passphrase:108

Fig. 10. Error handling using backspace.

The example presented in the figure is of the lower-bound keystroke values and corresponding upper-bound values for five users. User_a is granted access as long as the keystroke values provided at login fall within the range of 0.019 to 0.022, and this applies to other selected users as well. However, if user_c logs into the system as user_a, access is still granted because the keystroke times for user_a and user_c are similar. This method should always complement the usual UN-PW authentication method.

TABLE II. KEYSTROKE TIME VARIATION

User	User_a	User_b	User_c	User_d	User_e
Lower Bound	0.019	0.015	0.019	0.023	0.016
Upper Bound	0.022	0.020	0.023	0.026	0.019

In the error-handling approach, the last character typed and its corresponding keystroke are deleted once the backspace key is pressed. However, in rare cases, a user holds a character down for too long and causes it to duplicate, resulting in a key down time for each duplicated character but only one key up time for all of them. This arrangement renders the algorithm inefficient because the user's effort to delete the duplicated characters removes the data for other characters typed as well. Therefore, the user is required to delete all of the characters and start typing over again or refresh the page.

C. The Behavioral Dynamics Authentication Model

Five major constructs that directly influence the security of the authentication mechanism were developed based on the comprehensive review of the literature and were validated in the authentication experiment: the HCI devices, the encryption standard for the behavioral data, users' emotional state at the time of data entry, the seamless cross-platform transferability of the authentication mechanism, and the cost-effectiveness of the authentication mechanism (Fig. 11). Since the emergence of the keyboard, the HCI interface domain has advanced to support specialized devices that incorporate virtual or augmented reality and wearable technologies. Similarly, any authentication mechanism must take into account emerging trends in cryptology, including encryption algorithms. The accuracy of authentication depends on the emotional state of the authenticator as reflected in login errors. In other words, consideration of users' dominant emotional state during authentication can mitigate unauthorized authentication as well as authentication errors and recourse to the "forget your password?" option. For security of authentication mechanisms in a cross-platform domain, incorporating options for multiple operating systems authentications into the mechanism ensure seamless and secure operations. With the rapid application of AI in securing information system entities, and the emergence of innovative web applications taking into account the dynamic adoption of these into authentication mechanisms can ensure a balanced cost-security feature.

The system achieved 100% model validation with just five users by taking into account only one feature of each of the five constructs. This result demonstrates the economic feasibility and potential scalability of the model for future research involving experiments with various combinations of the constructs.

The main findings of the present study include:

- 1) The introduction of five constructs that offer an optimal combination ensuring efficient and effective behavioral authentication across all computer usage contexts.
- 2) The assessment of KD simulation to showcase the capability of KD in accurately authenticating even with just five users,
- 3) The demonstration of achieving a 0% rate for both FAR and FRR, and
- 4) The feasibility of encrypting behavioral data to enhance data protection.

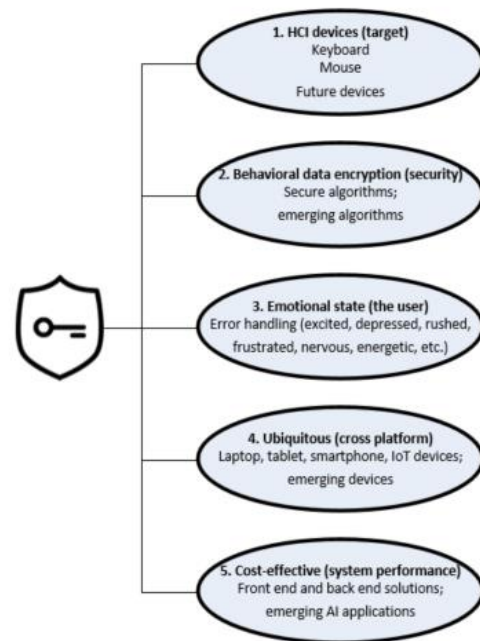


Fig. 11. The behavioral dynamics authentication model (BDAM).

V. CONCLUSION

Authentication processes that employ behavioral dynamics can ensure seamless and secure authentication through an optimal combination of HCI devices, data encryption, users' emotional states, cross-platform functionality, and the appropriate selection of solutions. We demonstrated the feasibility of constructing a secure, scalable, and dynamic behavioral authentication model, as described in this study. Our experimental results involved keystroke behavioral data collected from a computer system utilizing a web-based solution. We employed a CNN classifier and incorporated an error-correction feature. Remarkably, we achieved 100% accuracy in our authentication model with just five users. The web-based keystroke authentication system uniquely identifies users by integrating the UN-PW with their keyboard typing patterns, storing the keystroke data for each character typed independently, based on the bound value of the largest and smallest keystroke values recorded. For error handling, once the backspace key is triggered during typing, the keystroke data for the last character typed is deleted, and, when the keystroke data acquired during the process do not fall within the bounds established for a user, access to the system is denied; otherwise, access is granted.

The limitations of the research described here, also present avenues for future study to improve authentication systems. First, regarding the error-detection phase, the researchers considered only the backspace key, though some users also use the delete key, and taking this additional feature into account could enhance the speed of genuine authentication. Second, the system receives input in the form of mouse dynamics and touchscreen dynamics for laptops, tablets, and smartphones, but, while the focus here is on KD, the addition of touchscreen dynamics along with KD can add an additional security layer (potentially as an option) to touchscreen-enabled devices.

Third, changes in the environment and variation in the emotional state of the authenticator can render authentication challenging.

Future research can address the limitations by considering the following aspects. Firstly, researchers could replicate the study across various contexts, encompassing diverse environmental conditions similar to those encountered during authentication. Secondly, by accounting for users' cultural differences, extending the research to encompass subjects from multiple cultural backgrounds can enhance the robustness of the findings. Thirdly, given the substantial potential of ML in this domain, experimenting with different classifiers can yield those that ensure efficiency and effectiveness. Lastly, since the proposed model's versatility allows replication with diverse construct combinations and its scalability accommodates future trends in devices, security concerns, user behavior, and technologies, researchers could strive to pinpoint the optimal set of construct variables for enhancing authentication security and performance.

REFERENCES

- [1] Crawley, K.: A deep dive into the Verizon 2020 data breach investigations report. <https://spycioud.com/blog/a-deep-dive-into-the-verizon-2020-data-breach-investigations-report/> (2020). Accessed 22nd november 2022
- [2] Raab, C., Szekely, I.: Data protection authorities and information technology. *Comput. Law Secur. Rev.* 33, 421–433 (2017)
- [3] Jadhav, C., Kulkarni, S., Shelar, S., Shinde, K., Dharwadkar, N.V.: Biometric authentication using keystroke dynamics. Paper presented at the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). (2017)
- [4] Ben Fredj, O., Mihoub, A., Krichen, M., Cheikhrouhou, O., Derhab, A.: CyberSecurity attack prediction: a deep learning approach. Paper presented at the 13th International Conference on Security of Information and Networks. (2020)
- [5] Aldwairi, M., Aldhanhani, S.: Multi-factor authentication system. Paper presented at the 2017 International Conference on Research and Innovation in Computer Engineering and Computer Sciences (RICCES 2017), Malaysia Technical Scientist Association. (2017)
- [6] Pagar, V.R., Pise, R.G.: Password security mechanisms: comparative study. Paper presented at the International Conference on Research in Intelligent and Computing in Engineering (RICE 2017).
- [7] Shankland, S.: Two-factor authentication helps but isn't as secure as you might expect. Retrieved from <https://www.cnet.com/tech/services-and-software/two-factor-authentication-isnt-as-secure-as-you-might-expect-world-password-day/> (2017). Accessed 1st December, 2022
- [8] Baig, E.C.: Google will warn you when your passwords are too simple to guess and used too often. Retrieved from <https://techxplore.com/news/2019-10-google-passwords-simple.html> (2019). Accessed 7th january 2023
- [9] Verizon Inc.: Verizon data breach investigation report. Retrieved from <https://enterprise.verizon.com/resources/reports/dbir/> (2018). Accessed 3rd February, 2023
- [10] The Healthy Journal: How many people get hacked due to weak passwords? <https://www.thehealthyjournal.com/frequently-asked-questions/how-many-people-get-hacked-due-to-weak-passwords#:~:text=Passwords%20are%20often%20identified%20as,to%20practice%20good%20password%20hygiene> (2023). Accessed 3rd march 2023
- [11] Maynes, M.: One simple action you can take to prevent 99.9 percent of attacks on your accounts. <https://www.microsoft.com/en-us/security/blog/2019/08/20/one-simple-action-you-can-take-to-prevent-99-9-percent-of-account-attacks/> (2019). Accessed 3rd March 2023
- [12] Wang, D., He, D., Wang, P., Chu, C.-H.: Anonymous two-factor authentication in distributed systems: certain goals are beyond attainment. *IEEE Trans. on Dependable Secure Computing* 12, 428–442 (2014)
- [13] Golla, M., Ho, G., Lohmus, M., Pulluri, M., Redmiles, E.M.: Driving 2FA adoption at scale: optimizing two-factor authentication notification design patterns. Paper presented at the 30th USENIX Security Symposium. (2021)
- [14] Bhattasali, T., Saeed, K.: Two factor remote authentication in healthcare. Paper presented at the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI). (2014)
- [15] Sridhar, S.: Mitigating snoop-forge-replay attack by integrating text-based and language-based traits with the keystroke verification system. *Int. J. Sci. Eng. Res.* 5, 56–60. (2014)
- [16] Council of the EU and the European Council.: The general data protection regulation. Retrieved from <https://www.consilium.europa.eu/en/> (2022). Accessed 6th March, 2023
- [17] ElShekeil, S.A., Laoyookhong, S.: GDPR privacy by design. Ph. D. dissertation, master's thesis. Stolkholm University (2017)
- [18] Shekhawat, K., Bhatt, D.P.: A novel approach for user authentication using keystroke dynamics. *J. Discrete Math. Sci. and Cryptogr.* 25, 2015–2027. (2022)
- [19] Ibrahim, M., AbdelRaouf, H., Amin, K.M., Semary, N.: Keystroke dynamics based user authentication using Histogram Gradient Boosting. *Int. J. Computers and Inf.* 10, 36–53. (2023)
- [20] Chandok, R., Bhoir, V., Chinnaswamy, S.: Behavioural biometric authentication using keystroke features with machine learning. Paper presented at the 2022 IEEE 19th India Council International Conference (INDICON). (2022)
- [21] S. M. S. Ahmad, B. M. Ali, and W. A. W. Adnan.: Technical issues and challenges of biometric applications as access control tools of information security. *International Journal of Innovative Computing, Information and Control*, 8 (11), 7983-7999, (2012)
- [22] E. Pagnin and A. Mitroksotsa.: Privacy-preserving biometric authentication: challenges and directions. *Security and Communication Networks*, 2017, 1-9, (2017)
- [23] N. Memon.: How biometric authentication poses new challenges to our security and privacy [in the spotlight]. *IEEE Signal Processing Magazine*, 34(4), 196-194, (2017)
- [24] S. S. Harakannavar, P. C. Renukamurthy, and K. B. Raja.: Comprehensive study of biometric authentication systems, challenges and future trends. *International Journal of Advanced Networking and Applications*, 10(4), 3958-3968, (2019)
- [25] Z. Gao, W. Diao, Y. Huang, R. Xu, H. Lu, and J. Zhang.: Identity authentication based on keystroke dynamics for mobile device users. *Pattern Recognition Letters*, 148, 61-67, (2021)
- [26] Hassan, M.A., Shukur, Z.: A secure multi factor user authentication framework for electronic payment system. Paper presented at the 2021 Third International Cyber Resilience Conference (CRC). (2021)
- [27] Zaidan, D., Salem, A., Swidan, A., Saifan, R.: Factors affecting keystroke dynamics for verification data collecting and analysis. Paper presented at the 2017 Eighth International Conference on Information Technology (ICIT). (2017)
- [28] Kuka, E., Bahiti, R.: Information security management: password security issues. *Acad. J. Interdiscip. Stud.* 7(2), 43. (2018)
- [29] Hoonakker, P., Bornoe, N., Carayon, P.: Password authentication from a human factors perspective: results of a survey among end-users. Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting. (2009)
- [30] Liao, I.-E., Lee, C.-C., Hwang, M.-S.: A password authentication scheme over insecure networks. *J. Comput. and Syst. Sci.* 72, 727–740. (2006)
- [31] Alshanketi, F., Traore, I., Ahmed, A.A.: Improving performance and usability in mobile keystroke dynamic biometric authentication. Paper presented at the 2016 IEEE Security and Privacy Workshops (SPW). (2016)
- [32] De Cristofaro, E., Du, H., Freudiger, J., Norcie, G.: A comparative usability study of two-factor authentication. *arXiv preprint arXiv:1309.5344*. (2013)

- [33] Siti Rahayu, S., Guan, T.T., Yusof, R.: Enhanced authentication for web-based security using keystroke dynamics. *Int. J. Netw. Secur. and Its Appl. (IJNSA)* 12, 1–16. (2020)
- [34] Williamson, J., Curran, K.: The role of multi-factor authentication for modern day security. *Semicond. Sci. and Inf. Devices* 3(1), 16–23. (2021)
- [35] Gupta, A., Khanna, A., Jagetia, A., Sharma, D., Alekh, S., Choudhary, V.: Combining keystroke dynamics and face recognition for user verification. Paper presented at the 2015 IEEE 18th International Conference on Computational Science and Engineering. (2015)
- [36] Banerjee, S.P., Woodard, D.L.: Biometric authentication and identification using keystroke dynamics: a survey. *J. Pattern Recognit. Res.* 7, 116–139. (2012)
- [37] Oak, R.: A literature survey on authentication using behavioural biometric techniques. Paper presented at the Intelligent Computing and Information and Communication: Proceedings of 2nd International Conference, ICICC (2018)
- [38] Phadol, N.B.: Keystroke dynamics for user authentication using SCM. Master's thesis, National College of Ireland, Dublin. (2022)
- [39] Anusas-amornkul, T., Wangsuk, K.: A comparison of keystroke dynamics techniques for user authentication. Paper presented at the 2015 International Computer Science and Engineering Conference (ICSEC). (2015)
- [40] Quimatio, B.M.A., Njike, O.F.Y., Nkenlifack, M.: User authentication through keystroke dynamics based on ensemble learning approach. Paper presented at the CARI 2022-Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées. (2022)
- [41] Kar, S., Bamotra, A., Duvvuri, B., Mohanan, R.: KeyDetect: detection of anomalies and users based on keystroke dynamics. *arXiv preprint arXiv:2304.03958*. (2023)
- [42] Raul, N., Shankarmani, R., Joshi, P.: Non-conventional factors for keystroke dynamics as a support factor for authenticating users. *Int. J. Innov. Tech. and Exploring Eng. (IJITEE)*, 9, 474–479. (2020)
- [43] Shi, Y., Wang, X., Zheng, K., Cao, S.: User authentication method based on keystroke dynamics and mouse dynamics using HDA. *Multimed. Syst.* 1–16. (2022)
- [44] Wang, X., Shi, Y., Zheng, K., Zhang, Y., Hong, W., Cao, S.: User authentication method based on keystroke dynamics and mouse dynamics with scene-irrelevant features in hybrid scenes. *Sensors* 22, 6627. (2022)
- [45] Piugie, Y.B.W., Di Manno, J., Rosenberger, C., Charrier, C.: Keystroke dynamics based user authentication using deep learning neural networks. Paper presented at the 2022 International Conference on Cyberworlds (CW). (2022)
- [46] Mao, R., Wang, X., Ji, H.: ACBM: attention-based CNN and Bi-LSTM model for continuous identity authentication. Paper presented at the Journal of Physics Conference Series. (2022)
- [47] Stragapede, G., Delgado-Santos, P., Tolosana, R., Vera-Rodriguez, R., Guest, R., Morales, A.: TypeFormer: transformers for mobile keystroke biometrics. *arXiv preprint arXiv:2212.13075*. (2022)
- [48] Bhattacharya, P., Trivedi, C., Obaidat, M.S., Patel, K., Tanwar, S., Hsiao, K.-F.: BeHAuth: A KNN-based classification scheme for behavior-based authentication in Web 3.0. Paper presented at the 2022 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI). (2022)
- [49] Kang, S.J., Kim, S.K.: User interface-based repeated sequence detection method for authentication. *Intell. Autom. and Soft Computing* 35, 2573–2588. (2023)
- [50] Rahman, K.A., Neupane, D., Zaiter, A., Hossain, M.S.: Web user authentication using chosen word keystroke dynamics. Paper presented at the 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). (2019)
- [51] Vasyly, A., Sharapova, E., Ivanova, O., Denis, G., Yuliia, S.: Web-based application to collect and analyze users' data for keystroke biometric authentication. Paper presented at the 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). (2017).
- [52] Boakye Osei, M., Opanin Gyamfi, E., Okoe Alhassan, M.: Keystroke dynamics algorithm for securing web-based password driven systems. *Asian J. Res. in Comput. Sci.* 4, 1–26. (2020)
- [53] Aliekhsiev, V., Strelnitskiy, A., Gavva, D., Gorelov, D., Synytsia, Y.: Studying keystroke dynamics statistical properties for biometric user authentication. Paper presented at the 2018 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). (2018)
- [54] Saravanan, A., Bama, S. S.: CloudSec (3FA): a multifactor with dynamic click colour-based dynamic authentication for securing cloud environment. *Int. J. Inf. and Comput. Secur.* 20, 269–294. (2023)
- [55] Yang, H., Meng, X., Zhao, X., Wang, Y., Liu, Y., Kang, X., . . . Huang, W.: CKDAN: Content and keystroke dual attention networks with pre-trained models for continuous authentication. *Comput. and Secur.* 128, 103159. (2023)
- [56] Cockell, R., Halak, B.: On the design and analysis of a biometric authentication system using keystroke dynamics. *Cryptogr.*, 4(2), 12. (2020)
- [57] M. S. Sayeed, P. P. Min, and M. A. Bari.: Deep Learning Based Gait Recognition Using Convolutional Neural Network in the COVID-19 Pandemic. *Emerging Science Journal*, 6(5), 1086-1099, (2022)
- [58] M. W. A. El-Soud, T. Gaber, F. AlFayez, and M. M. Eltoukhy.: Implicit authentication method for smartphone users based on rank aggregation and random forest. *Alexandria Engineering Journal*, 60(1), 273-283, (2021)

Visualization of AI Systems in Virtual Reality: A Comprehensive Review

Medet Inkarbekov¹, Rosemary Monahan², Barak A. Pearlmutter³

Department of Computer Science, Maynooth University, Maynooth, Co Kildare, Ireland^{1,2,3}

Hamilton Institute, Maynooth University, Maynooth, Co Kildare, Ireland^{2,3}

Abstract—This study provides a comprehensive review of the utilization of Virtual Reality (VR) in the context of Human-Computer Interaction (HCI) for visualizing Artificial Intelligence (AI) systems. Drawing from 18 selected studies, the results illuminate a complex interplay of tools, methods, and approaches, notably the prominence of VR engines like Unreal Engine and Unity. However, despite these tools, a universal solution for effective AI visualization remains elusive, reflecting the unique strengths and limitations of each technique. The application of VR for AI visualization across multiple domains is observed, despite challenges such as high data complexity and cognitive load. Moreover, it briefly discusses the emerging ethical considerations pertaining to the broad integration of these technologies. Despite these challenges, the field shows significant potential, emphasizing the need for dedicated research efforts to unlock the full potential of these immersive technologies. This review, therefore, outlines a roadmap for future research, encouraging innovation in visualization techniques, addressing identified challenges, and considering the ethical implications of VR and AI convergence.

Keywords—Virtual Reality (VR); Artificial Intelligence (AI) Visualization; VR in AI Visualization; Human-Computer Interaction (HCI)

I. INTRODUCTION

Artificial intelligence (AI) systems have advanced quickly in recent years, with applications ranging from robotics and autonomous vehicles to machine learning and natural language processing [1]. As these systems become more complex, effective visualisation methods are becoming increasingly important [2]. Primarily, it fosters an understanding of AI operations, demystifying the 'black box' nature of these systems into interpretable processes [3]. This increased transparency aids not only in understanding the system's operations but also in spotting errors or inefficiencies, thus, paving the path for necessary improvements. Visualization also plays a pivotal role in enhancing the trust and transparency of AI systems, by illuminating the decision-making pathways of these technologically advanced systems [4]. Lastly, visualization is a key player in education [5], converting abstract AI concepts into tangible, clear forms. Thus, it stands as an indispensable tool in making intricate AI models more understandable, trustworthy, and user-friendly.

Virtual reality (VR), with its immersive and interactive capabilities, offers a promising platform for visualizing AI systems [6]. Users can experience and interact with AI systems in an immersive and interactive environment via VR, and visualization techniques can aid users in better grasping and analyzing the data produced by AI systems.

The real-world significance of comprehending AI systems

cannot be overstated. In sectors like healthcare, finance, and transportation, misinterpretations or biases stemming from obscure AI predictions can lead to severe consequences, including misdiagnoses, financial inaccuracies, and transportation mishaps [7, 8]. Proper visualization tools can mitigate these risks by illuminating the decision-making processes, ensuring more accurate and safer AI-driven outcomes. Moreover, a transparent understanding of AI systems can bolster trust among end-users, a crucial factor for the widespread adoption and societal acceptance of these technologies [9]. The economic implications are profound; AI systems that are both effective and trusted can revolutionize industries, driving innovation and efficiency [10].

The immersive experience of VR can offer new possibilities to make Convolutional Neural Networks (CNNs) more human-understandable by depicting them in 3D environments [11]. However, according to [12], this field is still relatively unexplored and has two main challenges. First, rendering large, popular architectures like ResNet50 is currently not feasible with existing tools, which often restrict CNNs to linear structures without splits or joints. Additionally, the number of visible layers may be limited due to computational limitations or constraints with the interaction design. Second, current tools do not offer a flexible and convenient interface for developers and researchers to visualize custom architectures. To fully realize the potential of VR for understanding and interacting with CNNs, more research and development are needed in this field. Despite the potential benefits of visualizing AI systems in VR, research in this area has been limited. While various studies have investigated the use of VR for visualizing AI systems, a comprehensive review or meta-analysis has yet to consolidate the present state of knowledge in this field.

This comprehensive literature study is intended to give an in-depth overview of the present state of knowledge about the viability of virtual reality technology for visualizing AI systems. We anticipate that the findings of this review will offer valuable insights that can guide future research in the domain of VR-based AI visualization. The review highlights the potential and limitations of existing VR engines and visualization methods, thereby identifying key areas of focus for further development and innovation. This study provides a basis for understanding the challenges currently faced in this field, such as handling complex data and managing cognitive load within VR environments. It also underscores the need for further exploration into the ethical implications of integrating VR and AI technologies. Thus, it presents a roadmap to shape the evolution of AI visualization tools, encouraging the development of more intuitive, user-friendly, and ethically

responsible solutions.

The main objectives of this review are to synthesize the findings from the literature, identify major themes and trends in the field, and offer suggestions for further investigation. In order to achieve these objectives, the following research questions are proposed:

- RQ1: How have different combinations of Virtual Reality techniques and AI systems been employed in the study of AI visualization in virtual reality, and which tools and frameworks have been most commonly utilized?
- RQ2: What are the diverse visualization and interaction techniques used for understanding, interpreting, and explaining complex AI models, and how do these methods enhance the exploration and representation of AI systems?
- RQ3: What are the potential applications, challenges, and technical considerations in utilizing VR technology for AI visualization, and how have these been addressed in various domains and studies?

By addressing these research questions, we aim to contribute to the existing knowledge on utilizing virtual reality technology in the visualization of AI systems and present an up-to-date overview of the relevant literature.

This paper is organized as follows: Section II begins with a comprehensive overview of the research methods employed. This encompasses the search strategy adopted, inclusion and exclusion criteria applied, study selection process, and methods of data extraction and synthesis, all focusing on the literature related to AI visualization in VR. Following this, Section III elaborates on the findings of the review, examining the array of VR techniques and AI systems in use, as well as the applied visualization and interaction techniques. It also discusses the potential applications and implementation challenges that VR technology faces in AI data visualization. Within this section, a discussion subsection synthesizes the literature, emphasizing the dynamic interplay of tools, methods, and approaches, and pointing out research gaps and challenges. Finally, Section IV concludes the paper, reflecting on the current state of the field and offering recommendations for future endeavors in VR and AI visualization.

II. METHODS

This comprehensive review examined research published in English in peer-reviewed journals or conference proceedings on the visualization of AI systems in VR but eliminated studies that still need to meet these requirements. By the type of AI system represented in VR, studies were categorized for synthesis in order to find common themes and trends and provide a thorough overview of the current state of the art.

A. Search Strategy

An exhaustive search was conducted in this comprehensive literature review to identify relevant studies on visualizing AI systems in virtual reality. The following databases were searched: IEEE Xplore, Scopus, Web of Science, Springer, and Google Scholar. The search strategy included relevant keywords and subject headings related to visualization techniques

and virtual reality AI systems,¹ and was limited to articles published in English, without a time restriction. Additionally, reference lists of identified articles were manually scanned for potential additional studies. The most recent results were obtained through a search performed on January 8, 2023. All identified articles were imported into Mendeley's reference management software for further analysis and screening.

B. Inclusion and Exclusion Criteria

Studies were included in this comprehensive review if they met the following criteria:

- published in a peer-reviewed journal or conference proceeding,
- focused on the visualization of AI systems in virtual reality, and
- provided empirical data, such as case studies, experiments, or user evaluations

while studies were excluded if they:

- were not published in English,
- focused solely on AI systems or virtual reality, but did not specifically address the visualization of AI systems within VR environments, or
- lacked sufficient methodological detail.

C. Study Selection

Fig. 1 presents a flow diagram illustrating the literature search and selection process of this comprehensive review. Guided by a well-defined research strategy, we aimed to ensure the exhaustiveness and relevance of the included studies. The process began with the identification of 940 records from an extensive search of multiple databases. The Publish or Perish tool was employed as an additional filter at this stage to refine search results and ensure the quality of the selected studies [13]. This tool enabled a more precise assessment of the articles, leading to the inclusion of studies that held significant influence in the field and strong relevance to the research questions. Consequently, 915 records were disregarded due to irrelevance or non-compliance with inclusion requirements.

The remaining 25 works underwent a full-text assessment to determine their suitability for the review. Subsequently, seven works were excluded for various reasons, including insufficient data, a lack of focus on visualization of AI systems in virtual reality, or other criteria-based factors. Ultimately, the final research included 18 papers, laying a robust foundation for synthesizing and assessing the state of knowledge on the visualization of AI systems in virtual reality.

¹The specific search query was: ("virtual reality" OR "VR") AND ("artificial intelligence" OR "AI") AND ("visualization" OR "visualisation" OR "display" OR "mapping" OR "interpret*" OR "explain*")

D. Data Extraction and Synthesis

We organized and summarized the included studies in detail in Table I (in Appendix) for analysis.

The synthesis process involved extracting key information from each study, such as reference, publication type, year of publication, VR engine/framework, features, and code availability. Each study is concisely outlined in the table, highlighting its primary goals and outcomes. The review's objective is to discern recurring patterns, shared aspects, and distinctions among the studies, thereby providing in-depth insight into the current state of knowledge regarding virtual reality's role in AI system visualization.

The table encompasses a diverse range of studies, including conference proceedings, journal articles, and tool development projects, all focusing on the visualization of AI systems in virtual reality. Most studies utilize widely-known VR engines, such as Unity and Unreal Engine, for frontend visualization, and backend machine learning frameworks like TensorFlow, PyTorch, and Caffe2. These studies explore features like network architecture, layer design, feature maps, and user interaction. Some also provide code availability, fostering further exploration and development by the research community.

By compiling information from these studies, the review presents a full overview of current research on the visualization of AI systems in virtual reality. It points out key advances and identifies areas requiring more research, ensuring a robust and in-depth understanding of the topic and laying the groundwork for future investigations.

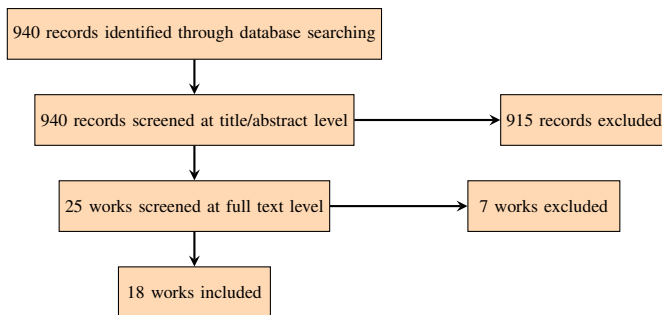


Fig. 1. Flow diagram of the literature search and selection process.

E. Study

The characteristics of each study, such as reference, publication type, year, features, VR engine/framework, and code availability, have been thoroughly analyzed and are presented in Table I. The table offers a comprehensive overview of the current research on the visualization of AI systems in virtual reality. By examining the characteristics of these studies, we aim to identify common themes, trends, and areas where further research is needed. This organized presentation of study characteristics in Table I allows for a clear understanding and comparison of the key aspects of the included works, facilitating a comprehensive synthesis of the existing knowledge in this field.

During the study selection process, we identified several studies that appeared to meet the inclusion criteria but were eventually excluded:

- 1) In [14], the authors excluded due to similarity to [11].
- 2) In [15], the authors Excluded due to significant overlap in content and authorship with [11].
- 3) In [16], the authors excluded due to its focus on visual data mining and representation of data and symbolic knowledge within VR spaces for classification tasks, rather than AI system visualization in VR.
- 4) In [17], the authors excluded because it focused on the simulation of autonomous and intelligent vehicles in virtual reality, rather than on the visualization of AI systems in VR.
- 5) In [18], the authors excluded as its main focus was on creating a virtual simulation environment for autonomous vehicles to learn obstacle avoidance using deep learning, rather than directly addressing the visualization of AI systems in VR.
- 6) In [19], the authors excluded as it involves virtual reality and deep learning, its primary focus is on pedestrian crossing behavior in the presence of automated vehicles rather than on the visualization of AI systems in VR.
- 7) In [20], the authors excluded as the paper is related to the broader field of AI and virtual reality; it does not specifically focus on the visualization of AI systems in VR. Instead, it presents a framework for detecting cybersickness in VR environments using machine learning models and explainable AI techniques. The goal of this research is to detect, analyze, and mitigate cybersickness in real-time for standalone VR headsets.

These studies were excluded to ensure a complete and unbiased perspective on the topic in our comprehensive review.

III. RESULTS OF SYNTHESSES

A. Virtual Reality and AI Systems (RQ1)

Different combinations of VR techniques and AI systems have been used in the study of AI visualization in virtual reality.

For example, [11] utilized the Unreal Engine for creating 3D scenes dedicated to neural network visualization. They used TensorFlow as their AI system to provide detailed information for neural networks and permit the adjustment of training parameters. Similarly, [12] employed the Unity game engine with OpenXR support for immersive visualization and interaction with convolutional neural networks (CNNs). Their AI system of choice was PyTorch, a deep learning framework designed to interconnect with Unity for dynamic visualization and interaction with custom CNN architectures.

Unity, as a gaming engine, was a common choice among researchers. For instance, [21] used Unity to create a virtual reality environment for visualizing deep neural networks built using the Caffe framework. The research [22] and [23] both utilized Unity in their research, illustrating its dual functionality. They used Unity for creating virtual reality environments and developing deep convolutional neural network models. Furthermore, they employed Unity Barracuda to define these neural network models, demonstrating that Unity can also serve as a host for AI systems. [24] paired Unity with the Oculus Quest VR headset and used TensorFlow and Keras

machine learning frameworks to define a CNN model for image classification.

The study [25] and [26] both used VR technology, with the latter also leveraging Unity and Steam VR, for deep learning model development. Their chosen AI system was deep learning neural networks defined using TensorFlow and Keras. It [27] and [28] also relied on Unity for virtual reality techniques, with the latter using TensorFlow as the AI system.

In some studies, the focus was primarily on the VR techniques utilized, with less emphasis placed on the specific AI systems implemented. For instance, [29] leveraged Open VR, which supports a range of consumer head-mounted displays (HMDs) including Oculus Rift and HTC Vive, yet they did not detail the AI systems used. Similarly, [30] employed an array of VR tools including an Oculus Rift headset, a Leap Motion '3D mouse', and a Microsoft Kinect sensor, but did not disclose the AI systems integrated with these tools. In another instance, [31] mentioned the use of explainable AI and clustering algorithms, but they did not specify the exact software libraries or frameworks employed, emphasizing the variance in the level of detail provided across different studies. In summary, a range of combinations of virtual reality techniques and AI systems are evident in the literature. Unity stands out as a common choice for VR, with TensorFlow being frequently paired as the AI system. This diversity underscores the flexibility and adaptability of these technologies in visualizing AI systems in virtual reality.

Beyond the specific visualization of AI systems in VR, broader AI applications in VR have also been extensively reviewed. For instance, [32] offers a comprehensive review of AI applications in VR, providing a wider perspective on how these two technologies can be combined to create interactive and immersive experiences.

B. Visualization and Interaction Techniques (RQ2)

Visualizing and interacting with complex AI models is crucial for understanding, interpreting, and explaining them. A variety of techniques are used to this effect, as evidenced by the following literature.

Interactive 3D visualization is a recurring technique used for exploring deep learning network layers at various levels of detail. The study [11] utilized this approach, enabling users to interact with each neuron in the network. Similarly, [12] used 3D rendering of Convolutional Neural Networks (CNNs), representing the computational graph as a connected conveyor system and optimizing the rendering of large architectures. They also allowed users to move, scale, and interact with CNN layers, offering a display of weight distributions, classification results, and feature visualizations. Expanding on these concepts, [33] delve into the merger of intricate networks and VR technology, creating interactive three-dimensional representations of large-scale datasets. Their approach is applied to a range of network types, including gene-gene interaction networks, social networks, and neural networks, demonstrating the broad applicability of these visualization techniques.

Visualization techniques can also involve more specialized approaches. For instance, [21] utilized interpretation modules, occlusion analysis, and virtual walkthroughs of network layers,

with real-time manipulation of input images for interaction. This study [24] employed 3D force-directed graphs and real-time color changes to represent neural network parameters, with direct manipulation and sonification for auditory feedback as interaction techniques.

Certain studies use more straightforward visual representations, such as [22], who rendered 2D images with the gray colormap using matplotlib, though interaction techniques were not explicitly discussed. The research [25] used a direct physical approach, visualizing interactions by moving concrete objects with hands and showcasing real-time test dataset results.

The research [26] and [27] presented network architecture, filters, and feature maps, with user interaction as the interaction technique. [34] visualized feature maps, filter responses, and saliency maps in 2D layers, while offering menu navigation, grabbing and moving objects, and selecting layer properties as interaction techniques.

The study [29] utilized an immersive node-link visualization based on VR, with neurons spatially arranged in circles, and interactive elements controlled through the spatial input devices included with the Head-Mounted Display (HMD). They also included a virtual travel technique (flying) for user positioning.

In cases such as [31], immersive parallel coordinates plots were used for visualization, although interaction techniques were not specified. The study [30] used immersive visualizations with SL Linden Scripting Language (LSL) and OpenSim, employing natural interaction techniques with commercial hardware.

Finally, [23] combined 3D visualization methods provided by TensorSpace.js and NeuralVis with conventional 2D visualization techniques such as Grad-CAM, while employing the Mixed Reality Toolkit and DXR for interaction.

In summary, a broad range of visualization and interaction techniques have been employed to illuminate and explore AI models. These methods offer different ways of understanding and interacting with AI systems, enriching our ability to interpret and utilize these complex models.

C. Application and Implementation (RQ3)

The utilization of VR technology for data visualization has vast potential, yet it is not devoid of challenges. For example, the complexity and high dimensionality of modern datasets can often make visualization difficult [6, 30]. Additionally, issues surrounding the scalability and adaptability of these visualization techniques to accommodate larger, more intricate architectures are prevalent [12].

Moreover, the cognitive load brought about by the complexity of the data and the VR environment itself is another concern. This has led to the call for more intuitive and user-friendly visualization tools to simplify the learning process and boost understanding [22, 24, 26, 27].

Moreover, the limitations of current tools and technologies present considerable technical challenges. For example, [22] pointed out the lack of support for large datasets in VR platforms and the limited resolution of existing VR display technology. The need for high-quality graphics and processing

power, along with the potential for motion sickness in VR environments, create substantial obstacles [35]. The study [12] say that in sensitive applications such as medicine and law enforcement, black-box systems such as neural networks will need to become more transparent in order to comply with regulations and earn the public's trust. This is compounded by the difficulties in understanding and representing the depth of images and models, as noted by [32]. These intersecting technical and ethical considerations necessitate careful attention and innovative problem-solving in the continued development and implementation of VR visualization tools for AI systems.

A recent study by [33] provides a practical example of VR-based model visualization in the context of gene-gene interactions between human sex chromosomes and other human chromosomes. The authors detail the entire process of model development, from data collection to the final display of the model in a VR environment. The study also outlines specific software and hardware requirements, including the use of Python libraries, Unity 3D software, and Meta Quest 2 hardware. Despite the technical challenges involved, the study underscores the value of VR technology in enhancing the visualization and interpretation of complex network data.

Notwithstanding these challenges, promising applications of VR for data visualization have been demonstrated across various domains. For instance, VR has proven effective in interpreting and understanding complex AI structures, such as deep learning models and convolutional neural networks [11, 12, 21, 26, 27]. The application of VR in these areas offers potential enhancements in explainability and interpretability of complex AI and machine learning systems [11, 21].

D. Discussion

The synthesis of the literature reveals a dynamic interplay [11, 12] of tools, methods, and approaches in utilizing VR for the visualization of AI systems. The widespread adoption of versatile VR engines like Unreal Engine and Unity [21–23] underlines the necessity for flexible and universally embraced foundations. These engines are compatible with an array of AI systems, including but not limited to ML platforms like TensorFlow, Caffe2, and PyTorch [24–26], showcasing their capacity to integrate with diverse technologies seamlessly.

The assortment of visualization and interaction techniques in the literature reflects the importance of intuitive and engaging methods for representing complex AI models. Interactive 3D visualization, immersive node-link visualization, and other specialized approaches illustrate the richness of techniques currently employed. However, the diversity of these methods underscores a potential research gap, as the lack of standardization could hinder collaborative efforts and delay the development of best practices.

Despite the widespread application of VR in AI visualization across various domains, significant challenges persist [35]. These include the high dimensionality and complexity of modern datasets, the cognitive load imposed by the VR environment, and the limitations of current tools and technologies. Thus, another research gap emerges: the need for more user-friendly, intuitive, and adaptable tools that can accommodate complex and high-dimensional AI models.

Moreover, the increasing integration of these technologies into our everyday lives necessitates a focus on ethical considerations. Several studies highlight the need for responsible and considerate use of AI and VR technologies, revealing another research gap that calls for more research into ethical guidelines or principles for this field.

Therefore, future research efforts must continue to innovate visualization and interaction techniques, address identified challenges, consider ethical implications, and fill the existing research gaps [33]. Moreover, it is crucial that subsequent studies continue exploring the potential of VR for AI visualization across diverse application domains, while seeking to reduce the cognitive load and enhance user experience.

IV. CONCLUSION

The application of Virtual Reality (VR) in the realm of Artificial Intelligence (AI) visualization represents a burgeoning frontier in Human-Computer Interaction (HCI). As this comprehensive review elucidates, the field is characterized by a rich interplay of tools, techniques, and methodologies, with VR engines like Unreal Engine and Unity at the forefront. However, the absence of a universal solution for AI visualization emphasizes the inherent complexities and challenges of this domain. The myriad of visualization techniques, while showcasing the field's innovative spirit, also points to a lack of standardization, potentially hampering collaborative advancements.

Challenges such as managing high data complexity and cognitive load, coupled with the limitations of current tools, remain significant hurdles. These challenges, however, also pave the way for future research opportunities, emphasizing the need for more intuitive and adaptable solutions.

In conclusion, the application of VR in AI visualization is an evolving field marked by significant potential and substantial challenges. The research reviewed here provides a solid foundation for future studies, pushing forward a deeper understanding and more effective use of AI systems through immersive technologies. Nonetheless, substantial gaps exist in the current research that need to be addressed to further advance this promising field.

FUNDING

This work is supported by Science Foundation Ireland, under Grant number 20/FFP-P/8853.

REFERENCES

- [1] S. A. Seshia, D. Sadigh, and S. S. Sastry, "Toward verified artificial intelligence." *Communications of the ACM*, vol. 65, no. 7, pp. 46–55, 2022.
- [2] A. Chatzimpampas, R. M. Martins, I. Jusufi, and A. Kerren, "A survey of surveys on the use of visualization for interpreting machine learning models," *Information Visualization*, vol. 19, no. 3, pp. 207–233, 2020.
- [3] Y. Liang, S. Li, C. Yan, M. Li, and C. Jiang, "Explaining the black-box model: A survey of local interpretation methods for deep neural networks," *Neurocomputing*, vol. 419, pp. 168–182, 2021.

- [4] P. Schmidt, F. Biessmann, and T. Teubner, "Transparency and trust in artificial intelligence systems," *Journal of Decision Systems*, vol. 29, no. 4, pp. 260–278, 2020.
- [5] W. Samek and K.-R. Müller, *Towards Explainable Artificial Intelligence*. Cham: Springer International Publishing, 2019, pp. 5–22. [Online]. Available: https://doi.org/10.1007/978-3-030-28954-6_1
- [6] E. H. Korkut and E. Surer, "Visualization in virtual reality: a systematic review," *Virtual Reality*, pp. 1–34, 2023.
- [7] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1721–1730.
- [8] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science (New York, N.Y.)*, vol. 366, no. 6464, pp. 447–453, 2019.
- [9] M. T. Ribeiro, S. Singh, and C. Guestrin, "'why should i trust you?': Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144.
- [10] A. F. Borges, F. J. Laurindo, M. M. Spínola, R. F. Gonçalves, and C. A. Mattos, "The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions," *International Journal of Information Management*, vol. 57, p. 102225, 2021.
- [11] A. Schreiber and M. Bock, "Visualization and exploration of deep learning networks in 3D and virtual reality," vol. 1033, 2019.
- [12] C. Linse, H. Alshazly, and T. Martinetz, "A walk in the black-box: 3D visualization of large neural networks in virtual reality," *Neural Computing and Applications*, vol. 34, no. 23, pp. 21 237–21 252, 2022.
- [13] A.-W. Harzing, "Publish or perish," Online, 2007. [Online]. Available: <https://harzing.com/resources/publish-or-perish>
- [14] M. Bock and A. Schreiber, "Visualization of neural networks in virtual reality using unreal engine," 2018.
- [15] A. Wohlan, N. Hochgeschwender, and N. Meissler, "Visualizing convolutional neural networks with virtual reality," *ACM*, 11 2019, pp. 1–2.
- [16] J. J. Valdés, E. Romero, and A. J. Barton, "Data and knowledge visualization with virtual reality spaces, neural networks and rough sets: Application to cancer and geophysical prospecting data," *Expert Systems with Applications*, vol. 39, 2012.
- [17] O. Lamotte, S. Galland, J.-M. Contet, and F. Gechter, "Submicroscopic and physics simulation of autonomous and intelligent vehicles in virtual reality," in *2010 Second International Conference on Advances in System Simulation*, 2010, pp. 28–33.
- [18] L. H. Meftah and R. Braham, "A virtual simulation environment using deep learning for autonomous vehicles obstacle avoidance," in *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2020, pp. 1–7.
- [19] A. Kalatian and B. Farooq, "Decoding pedestrian and automated vehicle interactions using immersive virtual reality and interpretable deep learning," *Transportation Research Part C: Emerging Technologies*, vol. 124, p. 102962, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0968090X2030855X>
- [20] R. K. Kundu, O. Y. Elsaid, P. Calyam, and K. A. Hoque, "VR-LENS: Super learning-based cybersickness detection and explainable AI-guided deployment in virtual reality," 2023.
- [21] A. Aamir, M. Tamosiunaite, and F. Wörgötter, "Caffe2unity: Immersive visualization and interpretation of deep neural networks," *Electronics (Switzerland)*, vol. 11, 2022.
- [22] L. Bibbò and F. C. Morabito, "Neural network design using a virtual reality platform," *Global Journal of Computer Science and Technology*, 2022.
- [23] H. Nagasaka and M. Izuhara, "Interactive visualization of deep learning models in an immersive environment," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '21. New York, NY, USA: Association for Computing Machinery, 2021.
- [24] Z. Lyu, J. Li, and B. Wang, "Alive: Interactive visualization and sonification of neural networks in virtual reality," 2021.
- [25] T. Zhang, "Research on environmental landscape design based on virtual reality technology and deep learning," *Microprocessors and Microsystems*, vol. 81, 2021.
- [26] N. Meissler, A. Wohlan, N. Hochgeschwender, and A. Schreiber, "Using visualization of convolutional neural networks in virtual reality for machine learning newcomers," 2019.
- [27] D. Queck, A. Wohlan, and A. Schreiber, "Neural network visualization in virtual reality: A use case analysis and implementation," in *Human Interface and the Management of Information: Visual and Information Design*, S. Yamamoto and H. Mori, Eds. Cham: Springer International Publishing, 2022, pp. 384–397.
- [28] T. Naraha, K. Akomoto, and I. E. Yairi, "Survey of the VR environment for deep learning model development," in *The 35th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI2021)*, Jun. 2021, p. 2N3IS2b04.
- [29] M. Bellgardt, C. Scheiderer, and T. W. Kuhlen, "An immersive node-link visualization of artificial neural networks for machine learning experts," in *2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020, pp. 33–36.
- [30] C. Donalek, S. G. Djorgovski, A. Cioc, A. Wang, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, S. Davidoff, J. S. Norris, and G. Longo, "Immersive and collaborative data visualization using virtual reality platforms," in *2014 IEEE International Conference on Big Data (Big Data)*, 2014, pp. 609–614.
- [31] S. Bobek, S. K. Tadeja, L. Struski, P. Stachura, T. Kipourous, J. Tabor, G. J. Nalepa, and P. O. Kristensson, "Virtual reality-based parallel coordinates plots enhanced with explainable AI and data-science analytics for decision-making processes," *Applied Sciences*, vol. 12,

- no. 1, p. 331, Dec 2021.
- [32] T. a. Ribeiro de Oliveira, B. Biancardi Rodrigues, M. Moura da Silva, R. Antonio N. Spinassé, G. Giesen Ludke, M. Ruy Soares Gaudio, G. Iglesias Rocha Gomes, L. Guio Cotini, D. da Silva Vargens, M. Queiroz Schmidt, R. Varejão Andreão, and M. Mestria, “Virtual reality solutions employing artificial intelligence methods: A systematic literature review,” *ACM Comput. Surv.*, vol. 55, no. 10, feb 2023. [Online]. Available: <https://doi.org/10.1145/3565020>
- [33] A. Hisham and B. Mahmood, “On the use of virtual reality in visualizing interactive network models for big data,” in *2022 8th International Conference on Contemporary Information Technology and Mathematics (ICC-ITM)*, 2022, pp. 20–24.
- [34] K. VanHorn and M. C. Çobanoğlu, “Democratizing AI in biomedical image classification using virtual reality,” *Virtual Reality*, vol. 26, no. 1, pp. 159–171, 2022.
- [35] G. Sharma, S. Chandra, S. Venkatraman, A. Mittal, and V. Singh, “Artificial neural network in virtual reality : A survey,” *International Journal of Virtual Reality*, vol. 15, 2016.

APPENDIX

A. Additional Data

TABLE I. INCLUDED STUDIES

Ref	Venue	Type	Year	VR Engine/Framework	Features	Code
[11]	Conference	Tool Development	2019	Unreal Engine/TensorFlow	Network architecture	—
<p>The paper describes an application aimed at enhancing the comprehensibility of neural networks through interactive 3D visualization. The app allows users to visualize the layers of a convolutional neural network and observe the classification process, providing greater transparency and opacity of AI systems for both experts and non-experts. The prototype will be improved based on feedback and evaluation, with plans to support additional model types and data formats, as well as integration with augmented reality headsets. Future enhancements will include the addition of user interfaces for displaying neuron results, and support for standard exchange formats such as ONNX.</p>						
[12]	Journal	Tool Development, Research	2022	Unity/ PyTorch	Network architecture, Layer design, Feature maps, User interaction	Available
<p>This work introduces an open-source software for the immersive visualization of popular CNN architectures using Python and the Unity game engine, allowing users to freely navigate a 3D environment in desktop or VR mode. The software offers feature maps, activation histograms, weight histograms, and feature visualizations to facilitate a greater understanding of CNNs. The authors address the issue of making the visualization of large-scale models feasible in VR by developing a Pytorch module that enables the optimized visualization of nearly any computational graph, including branches and joints, in Unity. This software is made for both experienced developers and researchers, as well as those who are new to the field of deep learning. In a use case study, the authors trained the architectures CovidResNet and CovidDenseNet using three distinct training strategies on the Caltech101 and SARS-CoV-2 datasets to produce models with varying generalization abilities. Using visualization software, the authors determined that CNNs memorized images based on high-frequency patterns and proposed new measures to make it more difficult for the network to memorize images. The ability to visualize popular, cutting-edge architectures raises new issues for future research, such as considering machine learning problems with 3D input data and visualizing statistical variations in network characteristics. The authors propose clustering the channels based on the cross-correlation of their filter outputs in order to further improve the presentation of the feature space and to illustrate the semantic connections between channels.</p>						
[35]	Journal	Survey	2016	—	—	—
<p>The article explores the potential of integrating ANNs with VR to create immersive and interactive experiences. The primary objective of the review is to investigate the possible benefits of combining ANN and VR, discussing various algorithms and training functions related to ANN and their role in solving VR development and interfacing problems. The authors address the importance of data visualization in VR, highlighting how VR can enable effective data interaction and manipulation. They discuss the use of ANN for creating VR spaces and dimensionality reduction techniques, such as clustering and neural networks, which can enhance the efficiency of VR systems. The paper also presents several applications of ANN in VR, including facial expression detection, human body tracking, face detection, data visualization, and speech recognition. The authors argue that the integration of ANN and VR can lead to the development of powerful systems capable of a wide range of applications, creating intelligent virtual environments. In conclusion, the authors emphasize the potential benefits of combining ANN and VR technologies, and how their synthesis can contribute to more responsive and stimulated tracking and analysis of events in various fields, ultimately transforming existing practices and conventions.</p>						
[21]	Journal	Tool Development	2022	Unity/ Caffe2	Network architecture, Layer design, Feature maps, User interaction	—
<p>The paper discusses the use of VR technology for visualizing and interpreting DNNs. The authors developed a plugin for the Caffe framework in the Unity gaming engine to create and visualize a VR-based AlexNet architecture for an image classification task. The interactive model allows users to navigate through the network and select connections to understand the activity flow of particular neurons. An interpretation module based on the Shapley values algorithm was used to analyze the network's decisions. The authors suggest that VR-based visualization can provide a more immersive and accessible way to explore and interpret DNN models, which can help improve their decision-making processes. They also suggest possible future work, including developing a formal quantification method for interpreting network decisions.</p>						
[22]	Journal	Research	2022	Unity/ Barracuda	—	—
<p>The article discusses using VR in deep learning and data visualization. The authors propose a VR platform using Unity for developing deep convolutional neural network models for image classification. The article describes the methodology used to create a CNN for recognizing human activities (HAR), which involves using the Barracuda package developed by Unity Labs and the ONNX format for transferring machine learning models. The article concludes that VR can be an effective tool for designing deep learning applications and is suitable for classifying images in various scientific sectors.</p>						
[24]	Conference	Tool Development	2021	Unity / Python	NN visualization: node-link approach, User interaction	—
<p>This study introduces a virtual reality interface for interacting with artificial intelligence. The interface lets users operate neural networks using virtual hands and offers audible feedback on the loss, accuracy, and hyperparameters in real time. The system's goal is to give a creative and user-friendly interface for interacting with AI, which may facilitate the understanding of the principles behind training neural networks. The study suggests several future directions, including enhancing the system's pedagogical benefits, looking at designs for larger neural networks, and experimenting with new forms of sonification to enhance the user experience.</p>						
[25]	Journal	Research	2021	—	—	—
<p>This paper presents a deep learning model development environment based on VR technology for image classification. The proposed DNN environment allows users to build neural networks by moving concrete objects with their hands, automatically transforming these configurations into a trainable model and providing real-time test dataset results. The study highlights the significance of interactive technology in addressing challenges in understanding and analyzing neural networks. The system aims to bridge the gap between professionals in different disciplines, offering a new perspective on the model analysis and data interaction. The results demonstrate that the proposed DNN method outperforms traditional PCA and SVM methods in classifying environmental landscape images. The paper also discusses the implementation of real-time image processing algorithms on FPGAs, emphasizing the advantages of using large memory and embedded multipliers.</p>						

TABLE I. INCLUDED STUDIES

Ref	Venue	Type	Year	VR Engine/Framework	Features	Code
[26]	Conference	Tool Development	2019	Unity and STEAM VR/ TensorFlow	Network architecture, Feature maps, User interaction	Available
<p>This paper investigates the potential of VR for visualizing CNNs for individuals who are new to machine learning. Because neural networks are so complicated, the authors present a VR-based visualization method to help people who aren't experts understand how CNNs work. The study emphasizes the role of virtual reality in creating an immersive and engaging environment that increases learning motivation while reducing distractions.</p> <p>An exploratory study was conducted with 14 participants, most of whom had little knowledge of CNNs. The results showed that the VR visualization method was easy to use, helped people learn, and made them more interested in learning about CNNs. Participants found the VR environment more comfortable, fun, and exciting than traditional desktop visualizations. Some participants noted that the immersive nature of VR helped them focus better on the complex architecture of CNNs.</p> <p>The authors suggest doing a follow-up study comparing how well VR visualization works with traditional desktop visualisation. This would help us learn more about how immersive environments affect how well people learn. The study also gives ideas for improving VR visualisation by letting users move and resize individual visualization elements to make the structure fit their needs and preferences. This research adds to the growing interest in using VR in schools and workplaces to help people learn and understand difficult ideas.</p>						
[27]	Conference	Tool Development	2022	Unity /—	Network architecture, Feature maps, User interaction	—
<p>The work aims to develop a VR application to visualize CNNs for machine learning beginners. The authors identified the need for an interactive and immersive visualization tool to aid in understanding CNNs, which can be complex and challenging for beginners. They conducted a proof-of-concept study with five participants and used the thinking-aloud method to evaluate the clarity of the VR environment. The study found that the VR environment was intuitive and helpful for users to understand the CNN components. However, some users found it difficult to understand the relationship between the input layer and feature maps. In addition, the visualization of the pooling layers did not stand out from the feature maps in their form of presentation. The authors plan to enhance the prototype based on the evaluation and conduct a user study to further evaluate its effectiveness.</p> <p>In conclusion, the study found that the VR environment was intuitive and helpful for users to understand the CNN components, but some areas need improvement. The authors plan to enhance the prototype and conduct further studies to evaluate its effectiveness. This work has important implications for machine learning, as it provides an interactive and immersive visualization tool to aid in understanding complex CNNs, which can lead to better learning outcomes for beginners.</p>						
[29]	Conference	Tool Development	2020	Open VR/ —	Network architecture, Feature maps, User interaction	—
<p>The article describes a new tool for visualizing ANNs using node-link diagrams in immersive virtual reality. The tool is targeted towards machine learning experts and was evaluated through an expert review. The results of the review showed that the tool was perceived as helpful in a professional context, which supports the hypothesis that node-link visualization can improve the workflow of machine learning experts. While more evaluation is needed, the authors are optimistic that their visualization tool could be actively used by experts in the field.</p>						
[28]	Conference	Research	2021	Unity/ TensorFlow	—	—
<p>This paper explores and proposes using visualization technology for deep learning and VR research projects, opening up new avenues for exploration in the field. The paper presents an outline of deep learning visualization research and case studies of deep learning visualization research using VR. The paper also identifies challenges that need to be addressed for effective visualization using VR, such as evaluation methods, high-quality operability, a wide range of customizability, scalability, and special support for beginners. Case studies and an overview of deep learning visualization research using virtual reality are presented in this paper.</p> <p>The research presented in this paper provides evidence that VR technology has the potential to enhance the way humans interact with deep learning models by providing insights into the processes of model construction and training. Significant difficulties are highlighted, and potential solutions are proposed, such as the use of new evaluation criteria and gamification to simplify deep learning for newcomers. Further, the paper argues that the proposed method's scalability and adaptability are crucial for experienced users and programmers. Overall, the paper shows the potential of virtual reality technology in deep learning research, and the authors suggest that implementing functions to customize filters and visualize results in real-time is essential for practical application.</p>						
[34]	Journal	Research	2022	Unity/ TensorFlow	Network architecture, Layer design, Feature maps, User interaction	Available
<p>In this study, the authors developed an interactive VR environment for constructing and analyzing deep learning models for biomedical image classification. The authors found that their proposed tool effectively enabled non-experts to understand and organize the structure of deep learning models and allowed users to build more accurate models and troubleshoot existing models faster when compared to a state-of-the-art drag-and-drop alternative. The authors also found that users enjoyed the experience in virtual reality significantly more, which they suggest can partially explain the higher objective scores, as positive emotions help with information retention. The authors argue that their interface achieved better outcomes through intuitive affordances for user actions, immersion, and ease of use.</p> <p>While introducing virtual reality presents challenges in educational and professional applications, such as cost, design, and physical limitations, the authors designed their VR environment to mitigate some risks. A more comprehensive VR platform could benefit expert use in the future, with improvements such as adding more layer types, developing new UI elements for fine-tuning layer properties, and enabling nonlinear model designs. The authors also suggest that improved visualization techniques and more explanatory elements could improve their tool's demonstrational benefit and interpretability. They conclude that future 3D/4D data visualization work can benefit from more straightforward navigation and a more accessible computational approach. Their proposed VR environment could be directly applied to cross-domain applications in biomedical image classification, providing sufficient benefits in understanding and prototype development.</p>						
[31]	Journal	—	2021	—	—	—
<p>The paper presents a refinement of the Immersive Parallel Coordinates Plots (IPCP) system for Virtual Reality (VR) and integrates data-science analytics, including explainable AI (XAI) methods, to enhance the visualization of multidimensional datasets in VR. The enhancements aim to automate part of the analytical work and assist users in pattern identification in complex datasets.</p> <p>The focus on visualization of multidimensional datasets and the integration of XAI methods in a VR environment aligns with the main focus of your systematic review, which is the visualization of AI systems in virtual reality.</p>						

TABLE I. INCLUDED STUDIES

Ref	Venue	Type	Year	VR Engine/Framework	Features	Code
[6]	Journal	Survey	2023	—	—	—
<p>The paper presents a systematic literature review on visualization in VR. It analyzes various techniques used in different domains and their collaboration. The review found that there is a growing body of research on immersive visualizations across various problem domains. However, only a few studies focus on creating standard guidelines for VR, and each study either provides an individual framework or relies on traditional 2D visualizations. Game engines are widely used, but they are not suitable for critical scientific studies. The paper mentions two examples of AI visualization in VR: Gradient-weighted Class Activation Mapping (GradCAM) for Deep Reinforcement Learning (DRL) algorithms and Caffe2Unity for visualizing neural networks. The review highlights the need for further research and alternative approaches to address design challenges, develop standard guidelines for VR, and ensure the accuracy and effectiveness of 3D visualizations.</p>						
[30]	Conference	Tool Development	2014	—	—	—
<p>The authors leverage commercial software development for virtual environments such as video games or virtual worlds and focus on developing scientific data visualization tools within such environments. They develop immersive visualizations of highly dimensional data sets using general-purpose visualization techniques and scripts. They propose using natural interaction techniques in immersive virtual reality with inexpensive commercially developed hardware.</p>						
[23]	Conference	Tool Development	2021	Unity/ Unity Barracuda	—	—
<p>The paper presents an interactive visualization system for DL models in an immersive environment. The immersive environment allows for unlimited displays and visualization of high-dimensional data, making it possible to analyze data propagation through layers and compare multiple performance metrics. The proposed system addresses the challenge of limited display area in desktop environments and aims to make the analysis of complex DL models more accessible for non-experts. The prototype system received positive feedback from machine learning engineers, but they viewed the visualization technology as a unique introduction to the immersive environment. Future work includes improving system design, evaluating usability, comparing performance with existing desktop systems, and exploring the benefits of immersive visualization in DL model analysis. The ultimate goal is to develop hybrid systems that complement existing tools rather than replacing them.</p>						
[33]	Conference	Tool Development	2022	Unity/ Python	Network manipulation, Real-time evaluation, User interaction	—
<p>This paper investigates the integration of complex networks and VR technology to create interactive 3D visualizations of large-scale data. This methodology is applied to various types of networks including gene-gene interaction networks, social networks, and neural networks. The authors illustrate the process of developing a VR-based visualization model using a biological dataset. They highlight both hardware and software requirements for implementing such VR visualizations. Examples from literature showcase the successful application of VR technology in understanding intricate relationships in these networks. The authors predict an increase in the use of VR technology for data visualization, particularly with the emergence of Metaverse concepts. As future work, they plan to develop a large-scale gene-gene interactions dataset and a VR interactive application to provide an efficient model for specialists to interact with large-scale data.</p>						
[32]	Journal	Survey	2023	—	—	—
<p>This comprehensive literature review investigates the application of AI methods in VR solutions, given the scarcity of such studies. The analysis involved locating and evaluating relevant documents from various databases, with a particular focus on AI's contributions to VR applications. The study observed that machine learning, specifically in the subfields of neural networks, deep learning, and fuzzy logic, is the most prevalent AI technique employed in VR. The application of AI in VR revealed multiple advantages, including high algorithmic efficiency and precision, especially in human-machine interaction and intelligent robotics. The study also revealed numerous real-world application fields such as emotion interaction, education, agriculture, transport, and health. However, the review highlighted several limitations, such as high computational cost and dataset dependence. This paper emphasizes the potential for future research focusing on finding new VR applications incorporating AI technologies, alongside a stronger emphasis on AR.</p>						

Symbol Detection in a Multi-class Dataset Based on Single Line Diagrams using Deep Learning Models

Hina Bhanbhro¹, Yew Kwang Hooi², Worapan Kusakunniran³, Zaira Hassan Amur⁴

Computer and Information Science Department, Universiti Teknologi,
PETRONAS Seri Iskandar, Perak Darul Ridzuan, Malaysia^{1, 2, 4}

Faculty of Information and Communication Technology, Mahidol University, Thailand³

Abstract—Single Line Diagrams (SLDs) are used in electrical power distribution systems. These diagrams are crucial to engineers during the installation, maintenance, and inspection phases. For the digital interpretation of these documents, deep learning-based object detection methods can be utilized. However, there is a lack of efforts made to digitize the SLDs using deep learning methods, which is due to the class-imbalance problem of these technical drawings. In this paper, a method to address this challenge is proposed. First, we use the latest variant of You Look Only Once (YOLO), YOLO v8 to localize and detect the symbols present in the single-line diagrams. Our experiments determine that the accuracy of symbol detection based on YOLO v8 is almost 95%, which is more satisfactory than its previous versions. Secondly, we use a synthetic dataset generated using multi-fake class generative adversarial network (MFCGAN) and create fake classes to cope with the class imbalance problem. The images generated using the GAN are then combined with the original images to create an augmented dataset, and YOLO v5 is used for the classification of the augmented dataset. The experiments reveal that the GAN model had the capability to learn properly from a small number of complex diagrams. The detection results show that the accuracy of YOLO v5 is more than 96.3%, which is higher than the YOLO v8 accuracy. After analyzing the experiment results, we might deduce that creating multiple fake classes improved the classification of engineering symbols in SLDs.

Keywords—Single line diagrams; engineering drawings; synthetic data; symbol detection; deep learning; augmented dataset

I. INTRODUCTION

An engineering drawing (ED) is an illustration of a schematic that demonstrates the operation or construction of an electrical system, procedure, or plant facility [1]. Engineering designs comprise of technical drawings such as mechanical or architectural blueprints, electrical circuits, and drawings [2]. In many different businesses, there is an increasing need for establishing digital systems for processing and analyzing these representations [3]. With such a framework, connected businesses will have the unusual opportunity to make extensive use of diagrams to direct their future practices.

A single-line diagram uses lines and symbols to represent the logical flow of power through physical processes and plant components. Although these components resemble each other in form and shape, they are highly asymmetrical in nature, which makes these documents complex [1]. Distinct power distributions are represented by lines of variable thickness, and each sign stands for a different component such as a

transformer, generator, motor, switch, etc. [4]. A typical SLD diagram may have over 50 different symbols, making it an information-rich visual representation. While placing a purchase order or even when project teams are scheduling their work, these drawings are carefully inspected in order to estimate the numbers of various pieces of equipment [5]. When symbols on SLD diagrams are functionally different but visually identical, as in Fig. 1, this process can become considerably more difficult and complex. As a result, distinguishing one symbol from another can be both crucial and difficult. Misreading or omitting any material can also cause severe internal disagreements and be damaging to the progress of a project.

Scientists, on the other hand, are looking into solutions for a power system to transform the conventional power system that existed before into an intelligent power system. The fusion of a power system with artificial intelligence is getting closer and closer as new technologies, like artificial intelligence, arise [6]. It is a common duty in modern businesses and academia to include artificial intelligence technology in power system dispatching software to speed up the process of creating circuit diagrams for power systems. A fundamental document in the power system, the principal wiring diagram of the power station is also commonly needed for viewing and change by the power system's dispatching users [7]. The current power dispatching system relies heavily on the work expertise of dispatchers for the creation and upkeep of station wiring diagrams, which not only raises the danger of safety mishaps in the power grid system but also drives up the cost of wiring diagram maintenance [8, 12]. Therefore, one difficulty facing the contemporary power sector was how to employ artificial intelligence technology to automatically build the station wiring diagram.

Generative models have also undergone significant progress and have been successfully used in numerous areas. One of those is the Generative Adversarial Networks (GAN), which has emerged as a well-known and frequently employed technique for producing content. Ian Goodfellow first introduced GANs in 2014 [9]. We will go over our GAN-based approach to solving the issue of imbalanced classes within the context of Methods section. Another difficult issue that affects a wide range of fields, including engineering drawings [10], is the under-or over-representation of one or more classes of symbols in the diagrams in the dataset [11].

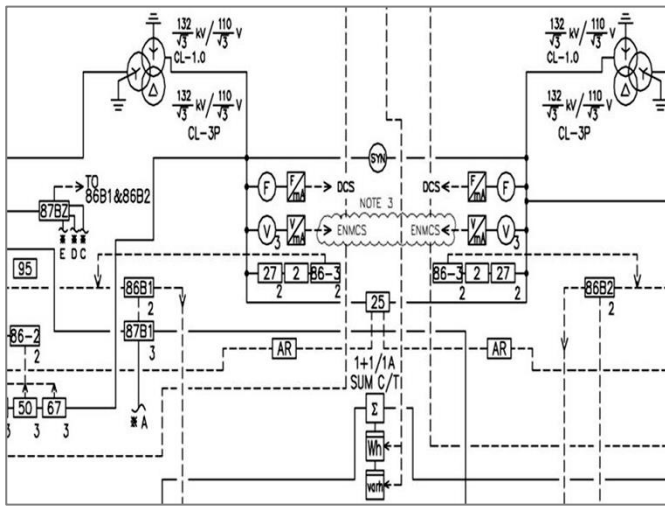


Fig. 1. A part of single line diagram.

It is logical to presume that construction industries have these designs for their on-going projects in a readable electronic format that can be edited with cutting-edge software. However, many businesses continue to maintain these designs as paper copies or in digitized form, particularly for their older projects. Therefore, digitizing these drawings in a way that makes information extraction simple and accessible, may be advantageous [13]. This can make it simple to correct previous designs when the plant's components have been replaced over time owing to maintenance. As a result, project teams will find it simpler to keep track of their instrumentation inventory during the building phase and to create a library of up-to-date drawings for maintenance during the post-installation phase with the help of digitized and updated SLD diagrams. The following restrictions are communicated through the contextualization and digitization of complicated SLDs:

A. Size

According to an estimate in [11], a typical SLD page consists of approximately 50 distinct types of shapes including symbols, connectors, and text. To depict a specific segment of a power system, it may be necessary to utilize anywhere between 100 to 500 pages.

B. Asymmetrical Components

Apart from the typical challenges of classical machine vision such as variations in lighting, scale, and pose, technical drawings utilize equipment symbols that conform to different standards across various industries. Consequently, assembling a precisely labeled dataset that can be employed for symbol classification is a complex undertaking, as mentioned in reference [14]. It is crucial to have a comprehensive assortment of precisely defined symbols that lack symmetry to effectively employ advanced deep learning methods for symbol recognition.

C. Connecting Lines

Connecting lines that indicate the logical and physical relationships between symbols are abundant and knotted in complex SLDs. As a result, it is difficult to apply digitization techniques based on thinning [15] or vectorizing [15]. The artwork for line identification are represented by lines of various

styles and thicknesses. Furthermore, sophisticated Engineering Drawings (EDs) adhere to rule sets for application-based connectivity. This means that based on a standard that cannot be stated or inferred from the use of the physical lines connecting the symbols, two symbols may or may not be connected. As a result, contextualization is more difficult to implement than when it is applied to simpler drawings, like circuit diagrams [16]. This opens up several intriguing options, such as incorporating human expert knowledge through human-machine interaction into a potential solution. Another avenue would be interactive learning [17].

D. Labels

Symbols, connectors, and other text characters may overlap; however, symbols and annotations in a variety of scripts and styles are used to distinguish between symbols exhibiting comparable characteristics, to indicate connectors, and to clarify additional information. Symbols with an overlap in drawing sheets are difficult to separate, as demonstrated by techniques like those used by Cao et al. [18] and Roy et al. [19]. Three further challenges have been identified once all of the text elements have been found: As seen in Fig. 1, various lengths and sizes are used to represent text strings that describe symbols and connectors. Additionally, it can be challenging to connect symbols and connectors to their matching text, and text interpretation mistakes could lead to some information being misunderstood.

E. Samples with Inconsistent Occurrence

Inconsistent appearance of symbols within the diagrams is another major issue towards digitization. Deep learning models perform better with large amounts of samples, while in SLDs, symbol frequency is highly imbalanced which creates a class imbalance problem due to the dominance of the majority classes over the minority classes. Hence, deep learning models can be biased towards the majority classes.

A range of methods, especially from the field of machine vision, must be applied to overcome these obstacles. These include symbol detection and localization, as well as feature extraction. The fact that recent advancements in deep learning and machine vision, particularly in the recognition and classification of objects, have not been put to the test against such challenging real-world situations, must be noted [21].

In this article, particularly, the YOLOv8 model for object identification and MFCGAN for class balancing are thoroughly examined. To extract symbols from drawing images, this study aims to use SLDs to create a dataset for model training. The dataset contains 22 different classes of symbols various shapes and sizes.

We were unable to locate a study that assesses a significant amount of deep learning-based detection algorithms that have been particularly designed for the problem domain of single-line symbol identification while taking into account key variables including Precision, Recall, and F1.

The following are the main contributions of this study:

- Symbols in SLD images are classified using YOLOv8, the latest variant of YOLO model.

- Mixed-quality single-line symbols are synthetically generated using MFC-GAN.
- A GAN-based solution is provided for enhancing the quantity of minority classes to handle the class imbalance problem; along with an expansion of the YOLOv5 training set using newly generated synthetic data.
- We suggested an experimental setup using MFC-GAN for creating synthetic images.
- The accuracy of symbol identification and recognition in Single Line Diagrams (SLDs) is enhanced by using a YOLOv5-based network for object detection.

According to experiments, the *IoU* and performance of the model can be enhanced through the use of synthetic image data generated using different GANs.

The remainder of this paper is structured as follows: In Section II, we delve into the landscape of existing research within the relevant domain, examining both the challenges that have been encountered and the solutions that researchers have put forth in this realm. Moving forward to Section III, we intricately explore the proposed methodology and perform an in-depth analysis of the dataset. Moreover, within this section, we provide a comprehensive exposition of the detection model. The outcomes of our dataset construction and symbol detection are meticulously presented in Section IV. Subsequently, we engage in a thorough discussion of the results in Section V. Lastly, we draw this study to a conclusion in Section VI.

II. RELATED WORK

This section covers recent accomplishments made by the research community in this domain. We discuss single-line engineering drawings, different deep learning techniques used for digitizing the engineering drawings, later we present GANs and discuss the general architecture and recent advancements made to improve the performance of GANs.

A. Single Line Diagrams

In various papers, including [1-4], the problem of recognizing and grouping symbols present in single-line diagrams (SLD) has been raised. The challenge of digitizing SLD, where the aim is to summarize the link between the numerous symbols, served as the inspiration for several of these works. The study in [22] provides an overview of numerous strategies created to digitize ED. In earlier research, including [23-26], symbols were recognized using classifiers that were traditionally based on machine learning and fed hand-crafted characteristics.

SLD digitization has notably drawn a lot of business interest due to the wide range of applications that may be made from a digital output, such as security evaluation, graphic simulations, or data analytics [27]. There are certain strategies developed expressly to handle the digitization of SLDs in the literature. More than 30 years ago, Furuta et al. [48] and Ishii et al. [28] published research on developing software to enable fully automated P&ID digitization. These techniques are currently ineffective due to incompatibility with hardware and software requirements. About ten years later, Howie et al. [29,

30] suggested a semi-automatic technique for localizing symbols of interest using the templates of the symbols as input. Gellaboina et al.'s [49] description of the most recent method for symbol identification uses an iterative learning strategy based on recurrent training of a neural network (NN) with the Hopfield model. This method was developed to pinpoint the most frequently occurring symbols in the artwork that also displayed a prototype pattern. Deep learning models were utilized [31] to build one-line diagrams automatically while generating core power systems.

B. Symbol Detection Using YOLO

Object detection can identify the sort of object present in an image or video and pin-point its location at the same time. In photos and videos, object detection expresses the location information as X and Y coordinate values. Additionally, the width and height values—which represent the object's size—are utilized as label information. Typically, the width and height data are expressed as bounding boxes using the X and Y coordinates.

Recent studies have employed deep neural networks to perform symbol spotting. For instance, researchers in [34] employed the YOLO 32, [33] model to identify symbols in floor plan diagrams. In another study [10], symbol detection was reformulated as a semantic segmentation problem, which led to the development of a pixel-level approach for symbol detection. Researchers are using YOLO for the goal of symbol recognition and classification as a result of the one-stage detection method's growing popularity and success [11]. To do this, the authors of [12] suggested transforming a construction image into a region adjacency network, where each node represented a connected component in the image. These nodes were then categorized using a YOLO. The YOLO and CNN-based technique was put forth in [13] and used to categorize symbols in [14].

Recent research has confirmed the effectiveness of YOLO variants in detecting complicated engineering components [35]. For instance, one-line symbols in substation diagrams were localized and categorized using YOLOv3. The model correctly identified 97% of the symbols. YOLO algorithms demonstrated encouraging results in detecting the symbols in electrical circuits despite the lack of suitable datasets [20]. Additionally, YOLO variations were used to accurately classify hand-drawn electric symbols with a 95% accuracy [11]. To the best of our knowledge, the work detailed here is the first attempt at localizing and matching symbols in a zero-shot method despite the very extensive literature that already exists in this field.

Since 2012, two primary types of deep learning-based object detection models have emerged: one-stage detectors and two-stage detectors, as described in research [32]. Understanding the concepts of region proposal and classification is essential to comprehend the distinction between the two categories. Region proposal refers to an algorithm that quickly identifies possible object locations, while classification is the process of categorizing objects based on their specific type. Although two-stage detectors are better at accurately detecting objects, their slow prediction time restricts their real-time detection ability. To address this issue,

one-stage detectors have been proposed that perform both classification and region proposal simultaneously, resulting in faster object detection. The one-stage detector is a technique that produces results by simultaneously executing classification and region proposal.

As depicted in Fig. 2, upon inputting the image to the model, the Convolutional Layer is employed to extract its features and perform classification. Simultaneously, a region proposal is conducted to generate the output. Models like YOLO, RetinaNet, RefineDet, etc. are good examples [37].

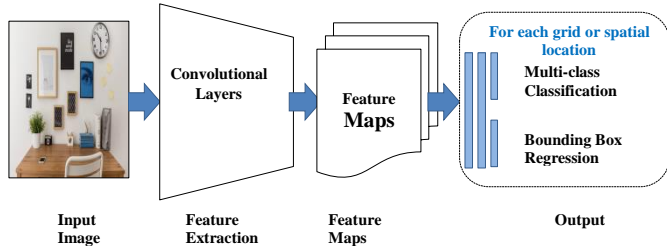


Fig. 2. One-stage model for object detection.

One of these one-stage detectors is YOLO, which integrates the region proposal and classification stages into a single operation. This means that it predicts the position and type of an object simultaneously by treating the bounding box and class probability as a single problem. YOLO divides the image into grids of a predetermined size to forecast the bounding box for each grid, and then trains the bounding-box confidence score and grid cell class score, as mentioned in reference [38].

The YOLO processing procedure is depicted in Fig. 2. First, an $S \times S$ grid area is created from the input image. The number of bounding boxes anticipated in each grid cell is equal to the number of bounding boxes that correspond to the area where an object is located. This can be denoted as (x, y, w, h) , where (x, y) denotes the center point coordinates of the bounding box, and (w, h) denote its width and height.

Second, the confidence, which stands for the box's dependability and is determined similarly to Equation (1). The IoU (Intersection over Union) is used to determine it by computing the ratio of the overlapping area between the predicted and ground truth bounding boxes divided by the probability $Pr(\text{Object})$, which represents the likelihood of an object being present in the grid.

$$Pr(\text{Object}) \times IoU^{\text{truth}} \quad (1)$$

The probability of C classes is then determined for each grid and Equation (2) is shown.

$$Pr(\text{Class}_i | \text{Object}) \quad (2)$$

In this instance, what is strange is that YOLO does not classify the number of classes (background) as an input to a neural network model, although the existing Object detection does [38]. YOLO divides the input image into grids in this manner, performing classification and bounding box calculations for each grid at the same time.

C. Synthetic Data Generation Using GANs

Several studies in the past decade have explored the challenge of identifying symbols in architectural floor plans. To overcome the scarcity of training data available for neural networks, the authors recommended employing a Generative Adversarial Network (GAN) to generate synthetic training data.

Ian Goodfellow first introduced generative adversarial networks (GAN) in 2014. (Goodfellow et al., 2014). These are regarded as generative models that can produce original content. The Generator (G) and the Discriminator (D) are two competing models (such as CNNs, neural networks, etc.) that make up GANs [39]. The discriminator is a classifier that gets input from both the generator and the training set (genuine content). (Fake input). The discriminator will learn how to differentiate between real input samples and bogus input samples during the training phase. However, the generator is trained to provide samples that accurately reflect the fundamental properties of the original data. (Replicating original content). The GAN model is shown in Fig. 3.

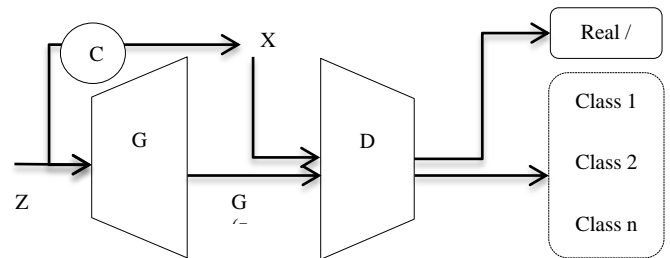


Fig. 3. Architecture of generative adversarial networks.

Equation (3) demonstrates that the value function is employed to perform adversarial training of both models G and D .

$$\min_D \max_G V(D, G) = E_{x \sim p_{\text{data}}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (3)$$

Where x is a sample from the real training data, $P_{\text{data}}(x)$ is the probability distribution over the real data, the probability distribution over the noise vector z is referred to as P_z , and the outcome of the generator function G (or generated images) is denoted as $G(z)$. GANs are at the forefront of image generation quality, as per [39].

GANs have been effectively used to solve a variety of issues, including speech synthesis, segmentation, and image production [40]. They have also been successfully used in recent years to address issues with class imbalance. The class mismatch is widespread throughout numerous industries, including banking, security, and health [6]. The issue arises when one or more classes are unequally or excessively represented in the dataset. When dealing with imbalanced datasets, a conventional supervised learning algorithm tends to favor the majority class [41].

By including conditional probabilities in the value function, supervised GANs offer an improvement over the basic GAN architecture. This gives the user more control over the samples that are created and introduces the diversity that is required to supplement synthetic input data for datasets with class

imbalance. Examples of this kind are AC-GAN [10], CGAN [9], and vanilla GAN [8]; even though the literature demonstrates that these models, particularly in extreme situations, can be significantly impacted by class disparity [11].

III. RESEARCH METHODOLOGIES

We present our method for recognizing the end-to-end symbols from intricate engineering drawings in Section III A. The dataset utilized for the tests will be covered in detail in the subsection that follows. Data exploration and pre-processing will be part of this. The specifics of our suggested approach to dealing with a class imbalance in these drawings are provided in Section IV.

Machine learning is commonly used to classify symbols and texts. Fig. 4 shows a conceptual model for digitizing engineering drawings that includes the essential phases. Such a framework will be extremely useful in fields where schematics can be turned into knowledge.

We determine the characteristics and variety of the created minority samples for our image-generating experiment after each run. In classification studies, we add created minority samples from trained models to the training data. (MFCGAN). The classification performances on the minority classes are then provided after a YOLO classifier has been trained on the expanded dataset.

A. Overview of Symbol Detection Framework

We first look for the areas of an engineering diagram that might contain interesting symbols and attempt to extract all the components from drawing. The next step is to locate and count the interesting symbols that originate from these zones of interest. The vast array of shapes and structures that these symbols emerge in drawings is the task's main problem. Furthermore, as stated in Section I, we cannot anticipate identical depictions of a specific component on all drawings. Additionally, there are a great number of different components and elements that are frequently used in these diagrams. As a result, it is not viable to use a fully supervised technique, training thoroughly to recognize and classifying every single type of object that could be seen in such images.

Information about the symbols that appear in an engineering drawing can be found in a variety of ways, including:

- 1) A table of legends listing the names of the components represented by the different symbols.
- 2) A table with numbers that represent the index of a component and the name of the object it represents.
- 3) There is no tabular data linking the names of objects to the appropriate diagrams.

In the current study, we focus on the first form of drawing, in which the component name and drawing image are both provided. We go into great detail on the various parts of the suggested framework in the sections that follow.

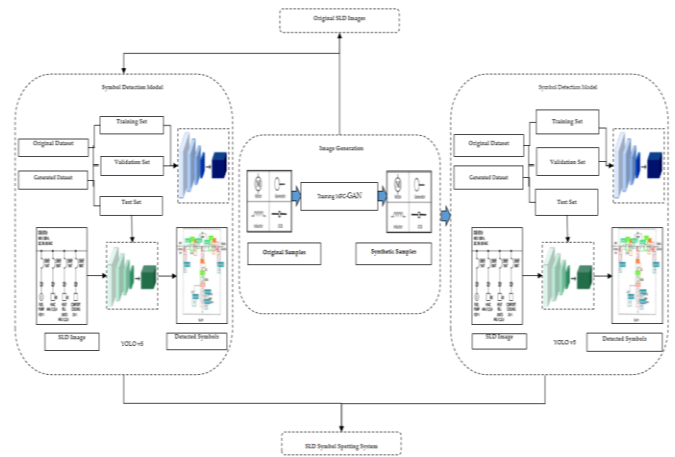


Fig. 4. Schematic of model for digitization of SLDs.

B. Summary of SLD Dataset

For the study in this paper, we chose to employ Single Line Diagrams (SLDs), as shown in Fig. 1. The engineering partner gave a set of 800 sheets for review. These diagrams contain a variety of symbols of varied sizes and dissimilar (asymmetrical) nature, as shown in Fig. 5.

The dataset is suitable for evaluation because the SLDs have a variety of attributes. The numerous electrical system components and connectivity information can be seen schematically represented on the SLD sheets. It is a representation of electrical apparatus and power flow movement, frequently in the form of symbols (represented as various kinds of lines).

In many industries, these diagrams can be found as paper documents or digital photographs. Evaluating and analyzing these materials requires a lot of experience, knowledge, and time [15]. Furthermore, misreading these publications can have disastrous repercussions. For instance, if an engineer needs to modify a wire in an electrical system after installation, they must first verify the associated SLD diagram and decide what safety precautions to take. Therefore, it's important to comprehend these designs correctly.

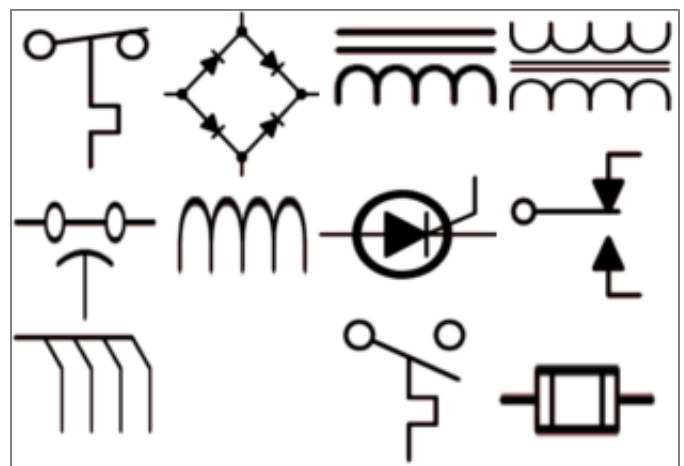


Fig. 5. Example of asymmetrical symbols found in SLDs.

The original data's big SLD sheets are 7500x5250 pixels in size. To expedite training, we divided the sheet into a 6x4 grid, resulting in 24 patches of sub-images that were minuscule in comparison to the original images (1250x1300).

The data generated by the annotation is kept in a file that corresponds to the 22 different classes. The width and height of the symbols that the bounding boxes enclose, as well as the center x and y values of the bounding boxes, were recorded as data. The collection of 800 images includes 12,500 samples, which represent 22 different types of symbols. The initial sample is severely imbalanced, as shown in Fig. 6.

A deep learning model needs to be fully annotated to be ready for training. To do this, we used the LabelBox program to annotate the set of SLD photos, as shown in Fig. 7. Twenty two different symbols in the total collection were annotated. Using the LabelBox tool to record the classes of the associated symbols and their locations is a simple approach for annotating a diagram.

In some instances, the distinctions between the symbols can be very significant. For instance, the dataset contains 1340 instances of generator symbols but only 99 and 117 instances of each disconnect and load symbol. Although delta and capacitor are present in the sample more than 800 times each, inductor and voltmeter are only present 203 and 212 times respectively. Three symbols that were significantly underrepresented overall were not included in the first trial (i.e. appears only once or twice in split sets).

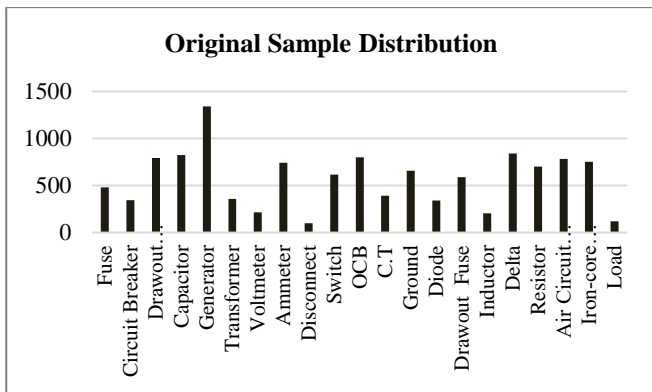


Fig. 6. Sample distribution in the original SLD dataset.

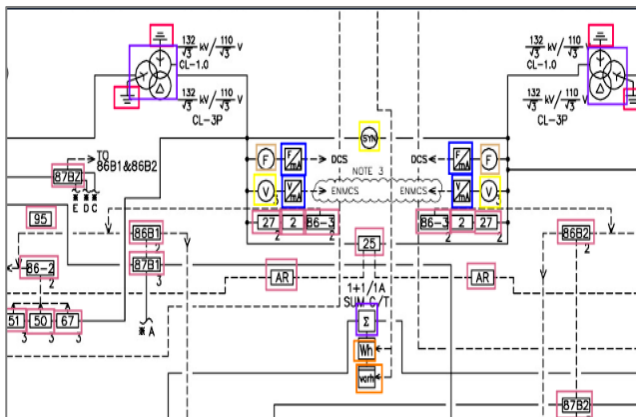


Fig. 7. An SLD annotated using labeling tool.

C. Symbol Detection

The YOLO approach was favored due to two key reasons. Firstly, it has a simple architecture that enables the prediction of multiple bounding boxes and class probabilities simultaneously using a single convolutional neural network. Secondly, YOLO is known for its high speed in comparison to other object detection techniques, which is essential for practical use in testing SLDs that contain an average of 50 engineering symbols.

1) *YOLOv8 architecture*: At the time this paper was being written, Ultralytics was actively working on YOLOv8 as they addressed community concerns and added new features. Glenn Jocher, the creator of YOLOv8, also discussed the developer-friendly features of YOLOv8 [54]. YOLOv8 comes with a CLI that enables training a model easier, in contrast to other models where chores are separated across numerous executable Python files. The addition of new convolutional layers and YOLOv8's Anchor Free Detection are further features of the software.

Since they may represent the distribution of the boxes from the target benchmark but not the distribution of the custom dataset, anchor boxes were a notoriously difficult component of older YOLO models. YOLOv8 is an anchor-free model, in contrast. In other words, rather than predicting an object's offset from a known anchor box, it predicts the object's center directly. To expedite Non-Maximum Suppression (NMS), a challenging post-processing procedure that sorts through candidate detection following inference, anchor-free detection decreases the number of box predictions [51, 55].

Using Equation (4), the bounding box's position is determined:

$$U_{x,y}^y P_{x,y} * IOU_{Ground Truth}^{Predicted} \quad (4)$$

According to Equation (4), x and y denote the yth bounding rectangle of the xth grid. The probability value assigned to the yth bounding box of the xth grid is $U_{x,y}$. If the yth bounding box contains an object, then $P_{x,y}$ is assigned a value of 1; otherwise, it is assigned a value of 0. The IoU between the predicted class and the actual ground truth is referred to as the $IoU_{groundtruth}$, and a greater IoU typically corresponds to more accurate predicted bounding boxes.

The bounding box, categorization, and confidence loss functions are combined to form the YOLOv8 loss function. The total loss function of the YOLOv8 is represented by Equation (5) [42]:

$$loss_{YOLOv8} = loss_{boundingbox} + loss_{classification} + loss_{confidence} \quad (5)$$

The stem's primary construction block, C2f, took the place of C3, and the first 6x6 conv is now a 3x3. Below is a diagram summarizing the module, where "f" represents the number of features, "e" represents the rate of growth, and CBS is a block made up of a conv, a BatchNorm, and a SiLU later. All of the

bottleneck's outputs are concatenated in C2f. C3 merely utilized the output of the previous bottleneck.

The first conv's kernel size was changed from 2x2 to 3x3, but the bottleneck remains the same as in YOLOv8. We might infer from this data that YOLOv8 is beginning to return to the ResNet block that was established in 2015. Features are directly concatenated in the neck without being forced to have the same channel dimensions. By doing this, the parameters count and tensor size as a whole are decreasing. YOLOv8 enhances photos while you're training online. The model views a slightly different variety of the images it has been given at each epoch.

2) *Multi-fake class generation*: Class imbalance has been a subject of extensive research, and various techniques have been developed, ranging from simple data augmentation and sampling to more sophisticated approaches like GAN [56]. In this study, we are utilizing MFC-GAN to generate more classes to handle the imbalance problem.

Our goal is to adopt a method similar to the MFC-GAN approach introduced in [57] to tackle the problem of class imbalance in the dataset of engineering symbols, specifically at the classification level.

The very little and occasionally subtle differences between the various classes of symbols led to the selection of this paradigm. We may train the discriminator using the MFC-GAN model to categorize both actual and false symbols, which allows for more precise discrimination across cases, as seen in Fig. 8.

By conditioning the generator on attribute labels, control generation was accomplished. Numerous studies involving various sample sizes in the minority classes, notably the goatee and eyeglass classes, were conducted. The MFC-GAN model is trained from scratch for each run, and samples are created following the end of the training.

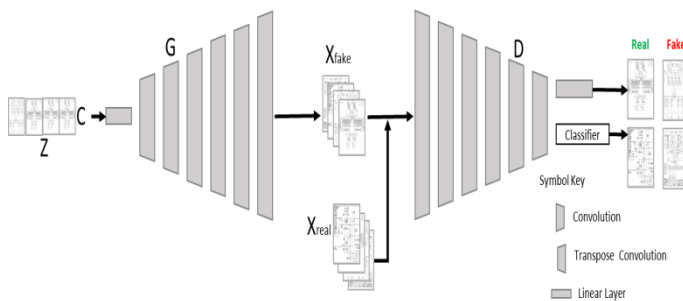


Fig. 8. Framework design for multi-fake class GAN.

The discriminator network for this study is built with four convolutional layers with two-stride spacing and uses batch normalization in between layers. Leaky ReLu with an alpha of 0.2 is used to activate all convolution layers, and the Sigmoid function is employed as the activation function in the final layer.

The classifier model generates a 2xN soft-max output for N classes and shares the discriminator layers with it. The generator is constructed using five transpose convolution layers

with a stride of two and one linear layer. All the layers except the last one are activated using Leaky ReLu, and the final layer is activated using a sigmoid function. Batch normalization is applied between adjacent layers.

The generator of the GAN model takes a noise vector with a size of 100 as input along with symbol label encoding, which is similar to the input of most GAN models. The label encoding is important for class-specific generation, which is a significant aspect of our experiment.

The generator produces a 64x64 image of greyscale symbols. A batch size of 100 and a learning rate of 0.001 were used, which were selected through experimentation. Both the discriminator and the generator employed spectral normalization. Eq. (6), (7), and (8) will be used to train the suggested model.

$$L_s = E[\log P(S=real|X_{real})] + E[\log P(S=fake|X_{fake})] \quad (6)$$

$$L_{cd} = E[\log P(C=c|X_{real})] + E[\log P(C' = c' |X_{fake})] \quad (7)$$

$$L_{cg} = E[\log P(C=c|X_{real})] + E[\log P(C=c|X_{fake})] \quad (8)$$

Where L_s denotes the chance that the sample is real or fraudulent and is used to determine the sampling loss. The losses for classification of both the generator and discriminator are calculated using L_{cd} and L_{cg} . The set of created images is called X_{fake} , and X_{real} represents the training data.

3) *Architecture of YOLOv5*: In this study, the YOLOv5 algorithm was utilized, which is one of the most recent variations of the YOLO algorithm [44]. This algorithm is a speedy and effective system for identifying objects and locating them instantly. Since the symbols present in SLDs have a high degree of similarity, rapid detection is also necessary, which the YOLOv5 algorithm can fulfill. The system was built using the PyTorch deep learning framework, which has excellent detection performance and has simplified the process of training and testing specialized datasets. The YOLOv5 algorithm comprises three components: the head, the neck, and the backbone [45].

For our investigation, we opted to utilize the YOLOv5 detection model because of its straightforwardness and transparency. YOLOv5 created CSPDarknet, which formed the core of the network [58], by combining Darknet with the cross-stage partial network (CSPNet) [43]. CSPNet addresses the issue of recurrent gradient information in large-scale backbones by integrating gradient changes into the feature map, which decreases the model's parameters and FLOPS (floating-point operations per second). This ensures inference speed and accuracy while also reducing model size, which is crucial for accurate and speedy recognition of sperm cells. Furthermore, the YOLOv5 incorporates a path aggregation network (PANet) [59] as its neck to improve information flow. PANet employs a novel feature pyramid network (FPN) architecture with an enhanced bottom-up methodology to increase low-level feature propagation. Adaptive feature sharing connects the feature grid to each feature level, ensuring that the downstream subnetwork receives meaningful data from every feature level. In addition, PANet enhances precise

localization signals at lower levels, considerably improving the object's location accuracy. The head of YOLOv5, the YOLO layer, generates three different sizes of feature maps to enable multi-scale prediction, allowing the YOLO model to handle small, medium, and large objects [58].

The CSPNet provides the framework for this algorithm. Because of the simplified model of CSPNet, fewer hyperparameters and FLOPS are produced, and the disappearing and ballooning gradient issues caused by complex neural networks are addressed. These enhancements improve the effectiveness and accuracy of object recognition inference. CSPNet has various features, including multiple convolutional layers, three convolutions in four CSP blocks, and spatial pyramid sharing. The CSPNet is responsible for extracting features from an input image, pooling and convolving that data to create a feature map. Consequently, in YOLOv5, the backbone serves as a feature generator [60].

The neck or core segment of YOLOv5 is referred to as the PANet. Its main function is to collect all the features obtained from the backbone, maintain them, and send them to the deeper layers to perform feature fusions. These feature fusions are then passed on to the head for object recognition, allowing the output layer to be aware of the high-level characteristics.

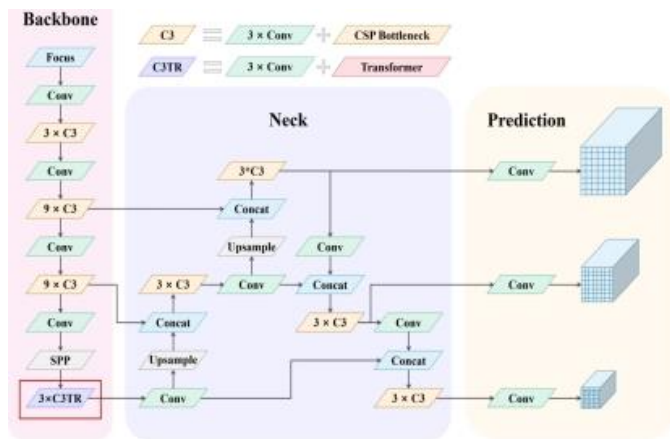


Fig. 9. Network architecture of YOLOv5 model.

The YOLOv5's head is responsible for identifying objects. It places bounding boxes and a class probability score around the target item, which is determined by 1x1 convolutions. The overall architecture of YOLOv5 is depicted in Fig. 9.

The position of the bounding box is established using Equation (9):

$$U_x^y = P_{x,y} * IOU_{Predicted}^{Ground Truth} \quad (9)$$

Equation (9) demonstrates that the yth bounding box of the xth grid is defined by x and y. The yth bounding box of the xth grid has probability value. $P_{x,y}$ equals 1 when a subject is present within yth bounding box; otherwise, it is equal to 0. The $IOU_{Ground Truth}$ is the IoU that exists among the predicted class and the actual data. Higher IoUs are related to more accurate predicted bounding boxes.

The loss function of YOLOv5 is produced by merging the bounding box, classification, and confidence loss functions.

The combined loss function of YOLOv5 is shown in Equation (10) [46].

$$loss_{YOLOv5} = loss_{bounding\ box} + loss_{classification} + loss_{confidence} \quad (10)$$

Equation (11) is used to determine the $loss_{bounding\ box}$.

$$loss_{bounding\ box} = \lambda_{if} \sum_{a=0}^{b^2} \sum_{c=0}^{d^2} E_{a,c}^g, h_g(2-k_a k_{na}) [(x_a - x'_a)^2 + (y_a - y'_a)^2 + (w_a - w'_a)^2 + (h_a - h'_a)^2] \quad (11)$$

Equation (11) uses h' and w' to denote the width and height of the target item, while x_a and y_a denote the coordinates of the target object in an image. Lastly, the indicator function (λ_{if}) shows whether the bounding box contains the target object.

The $loss_{classification}$ method is shown in equation (12):

$$loss_{classification} = \lambda_{classification} \sum_{c=0}^{b^2} \sum_{c=0}^{d^2} E_{a,c}^g \sum_{CCc1} L_a(c) \log(LL_a(c)) \quad (12)$$

$loss_{confidence}$ is determined using Equation (13):

$$loss_{confidence} = \lambda_{confidence} \sum_{c=0}^{b^2} \sum_{c=0}^{d^2} E_{a,c}^{Confidence} (c_i - c_1)^2 + \lambda_g \sum_{c=0}^{b^2} \sum_{c=0}^{d^2} E_{a,c}^{Confidence} (c_i - c_1)^2 \quad (13)$$

In Equations (12) and (13) show the symbols that represent confidence and signify the category loss coefficient λ , classification loss coefficient, and confidence score.

IV. RESULTS AND EXPERIMENTS

This section can be separated into two experiments and should present a clear and accurate depiction of the experimental findings, their analysis, and the conclusions that can be drawn from the experiments.

There were two experiments done. The initial test was created to assess a complete method for identifying symbols in engineering drawings. In this context, the aim is to enhance the overall efficiency of analyzing a collection of drawings by detecting and identifying symbols, which is an important task as symbols make up a significant portion of these drawings. This can aid in completing other tasks, such as detecting text, pipelines, etc. The second experiment is different from the first, as it concentrates on using GAN-based methods to deal with the problem of class imbalance.

A. Symbol Detection Using YOLOv8

The SLD sheets in our dataset had a size of about 7500x5250 pixels. To avoid using computationally expensive training data, the SLDs were divided into 24 patches by multiplying their original width by 6 and their original height by 4. The resulting patch size was approximately 1250x1300 pixels. The annotations for the entire SLD were used to retrieve data for each patch's annotations, as described in the preceding section.

Symbols that spanned multiple patches were excluded from the training phase. After extensive testing, the third version of the YOLO framework was selected as it showed better detection rates for small objects compared to the previous versions. It should be emphasized that, when compared to the

entire image size, the technical symbols in our dataset are relatively small.

To conduct the experiment, the researchers utilized a recent version of YOLO architecture. Initially, they configured the total number of classes to be 22 in all three YOLO layers, and then adjusted the number of filters to 3 (referred to as Class_{no} 5), where Class_{no} represents the complete number of classes present in the dataset.

The dataset was divided into two sets: training set with 640 SLDs and test set with 160 SLDs, with a ratio of approximately 80:20. A pre-trained YOLO network was used and fine-tuned on our dataset by adjusting all layers. YOLO was implemented using PyTorch. To enhance object detection for various object sizes, it was observed that changing the input size during training is effective [47]. In this study, the network input size was modified to 416x416 after every 10 batches, and the training stopped after 10,000 batches. The learning rate was 0.001, and the batch size was 64.

During the testing phase, the model input size was increased from 416x416 to 2400x2400. This enabled us to perform symbol detection on the original SLD images instead of integrating detection from the SLD patches, thereby simplifying the symbol detection process for an entire SLD diagram in one step. To evaluate the model, we experimentally set the Intersection over Union (IoU) threshold to 0.5 and compared the detected symbols with the ground truth. A Python-based front-end was developed utilizing OpenCV and other libraries to analyze and display manual errors.

B. Training Evaluation of YOLOv8

1) *Computer hardware configuration:* GPU computing is a preferred choice for processing deep learning on a PC [50], and therefore, strong hardware support is required for deep learning networks. The training and generation processes were conducted on a GPU workstation that ran on Linux, CUDA 11.1, Python 3.8, and PyTorch 1.8.0, and was equipped with an Nvidia A40 4 48 GB GPU.

C. YOLOv8 Detection Results

The training phase produced an accuracy of 96%, while the testing phase produced an accuracy of 95.9%, with 11987 out of 12500 symbols in the test set correctly detected and recognized. The loss matrix for the training and validation sets indicates that the most of the instances of the classes were identified and detected accurately, indicating that symbols with sufficient training instances were correctly identified. An example output from the proposed methods is shown in Fig. 10, with different symbols highlighted in different colors.

In this case, the identified symbols were labeled with numbers, and the labels predicted by the models were noted down to compare them with the actual labels later. These symbols comprised various electrical components like switches, generators, motor, re-lays, inductors, as well as input/output labels such as "label_to" and "label_from".

Table I in the paper contains details about the number of symbols in both the training and testing datasets, with columns labeled "No. of Training Samples" and "No. of Testing

Samples". It also displays the accuracy achieved for each class in the testing set, along with the number of symbols that were correctly identified, listed under "Correctly Detected Samples" or "Class Accuracy".

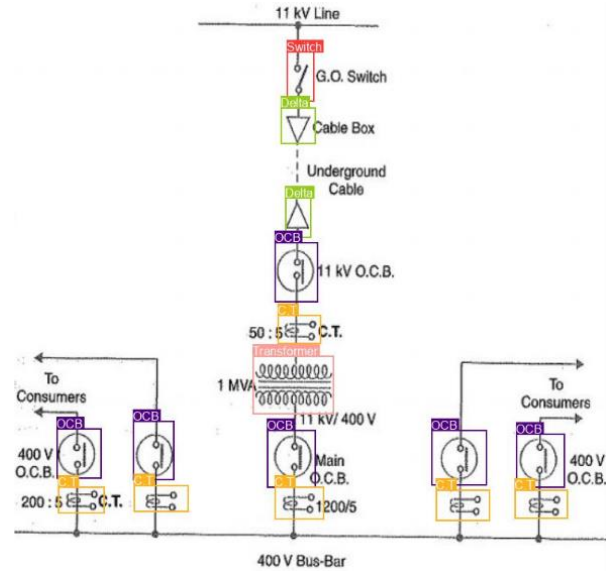


Fig. 10. Symbol detection using YOLOv8.

TABLE I. YOLOV8 DETECTION RESULTS BASED ON ORIGINAL SLD DATASET

Symbols	No. of Training Samples	No. of Testing Samples	Class Accuracy
Fuse	400	80	100%
Circuit Breaker	300	52	100%
Drawout Circuit Breaker	600	190	99%
Capacitor	601	20	99%
Generator	940	400	100%
Transformer	305	50	97%
Voltmeter	182	20	94%
Ammeter	550	190	100%
Disconnect	89	10	94%
Switch	501	115	100%
OCB	509	290	100%
C.T	309	80	100%
Ground	515	140	100%
Diode	310	30	98%
Drawout Fuse	679	110	100%
Inductor	185	18	98%

The results of the study show that most instances (11987 out of 12500) were accurately detected and identified. Fig. 10 displays different symbols from various SLD diagrams and how they can be accurately recognized regardless of their orientation. Some symbols, such as transformers, OCB, C.T, ground, and delta components, have various orientations, but the suggested approach can correctly detect and identify them. Even in situations where the text overlaps, resistors and ground

symbols are accurately detected, demonstrating the approach's resilience to innate visual issues, at least in this context, as opposed to conventional methods.

The round-shaped motor symbol in the testing set was misclassified as a generator symbol in all five instances due to their similarity. It is expected that increasing the number of training examples for this symbol, as well as other symbols in the majority class (such as switch, relay, C.T, inductor, voltmeter, load, etc.), would improve their detection rates.

To conclude, based on the results presented in Table I, it can be inferred that the detection rate for the symbols load, resistor, and motor was very low. This could be attributed to the fact that there were only a few training samples for these symbols. The complexity of the problem notwithstanding, the average accuracy of the remaining 22 symbols in the dataset was above 95%, which is a positive result, although these symbols were excluded.

D. MFC-GAN for Sample Generation and Symbol Detection

The purpose of this study is to evaluate a GAN-based model that can tackle the class imbalance problem in the dataset, which is a classification problem as opposed to a recognition task like in the first experiment. The main aim of this experiment is to utilize the MFC-GAN model to generate more symbols, which can then be added to the training set to improve classification accuracy.

Experiment 2 made use of a dataset that was very similar to the one utilized in Experiment 1, with all symbols being resized to 64x64 grayscale images. The problem was framed as a supervised learning task, where the goal was to learn a function $f(x)$ that maps a given engineering symbol instance (x_i) to its corresponding class (y_i). In this case, the 22 symbols in the dataset were represented as a discrete set of classes denoted as Y , where y_i belongs to Y . The dataset suffered from severe class imbalance, as previously mentioned, with some instances such as the angle choke valve being present in less than 0.01% of the dataset.

In this experiment, the MFC-GAN model was employed in two stages, namely the GAN training stage and the classification stage for this experiment. The purpose was to address the issue of class imbalance in the dataset. The MFC-GAN model was initially trained using all the samples in the dataset, with a focus on the symbols with the least representation. These symbols included fuse, circuit breaker, drawout circuit breaker, capacitor, generator, transformer, voltmeter, ammeter, disconnect, and switch. The numbers of occurrences of these symbols in the training set were 48, 344, 790, 1340, 821, 355, 212, 740, 99, and 616 respectively. The model underwent training only once on this dataset, and the generated samples were obtained after the training. To improve the learning of minority instance structure while training, the less represented classes were resampled.

Using the MFC-GAN model trained on the least represented symbols in the dataset, symbols from the minority class were generated. Eight symbols with the least representation were selected. To create a balanced dataset, 80% of the original dataset was used for training, and the remaining 20% was kept for testing. The artificially generated symbols

were added to the training set, providing more than 4,000 additional synthetic samples for each minority class. This allowed the dataset to be rebalanced by increasing the prevalence of the least represented symbols.

Our goal is to evaluate the effectiveness of the synthesized symbols by comparing the performance of a classification model trained before and after the inclusion of these symbols in the training set.

The experiment employed a four-layer CNN classification model, consisting of three convolution layers with 32, 64, and 128 outputs, respectively. The kernel sizes for these layers were 3x3, 2x2, and max-pooling. The fourth layer was a fully connected layer with 256 units, which represented the 22 symbol classes, and fed into a 22-way Soft-max out-put. The model was trained using SGD with 64 batches and a learning rate of 0.001. The model's classification performance was assessed using standard measures like true positive rate, balanced accuracy, G-mean, and F1-Score, with the aim of comparing the model's performance before and after incorporating the generated symbols into the training dataset.

1) Results: Fig. 11 displays a comparison between the original symbols in the diagram and the symbols generated by the MFC-GAN model.

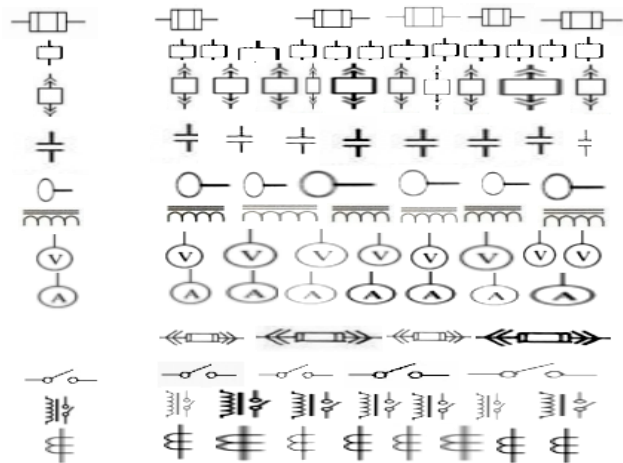


Fig. 11. Original SLD samples compared with MFC-GAN generated samples.

The comparison between the MFC-GAN generated symbols and the original symbols of the diagram is depicted in Fig. 11. The generated symbols by MFC-GAN were found to be more accurate and precise compared to symbols generated by other methods. The generated samples had clear symbol traits and distinct categories formed in each instance. Moreover, these high-quality samples resulted in an improved performance of the classifier. For example, Table III shows that for the angle disconnect, which had only 99 instances of the class, the accuracy improved from 0 to 94%. Similarly, seven out of eight minority classes demonstrated similar improvements. However, the MFC-GAN model did not improve the baseline in the load and inductor classes. It was observed that certain symbols such as OCB, capacitor, voltmeter, and ammeter exhibited significant similarity despite being uniquely generated, which hindered the classifier's ability to classify the load and inductor classes. This

observation was further supported by the low precision data in Table III for these classes.

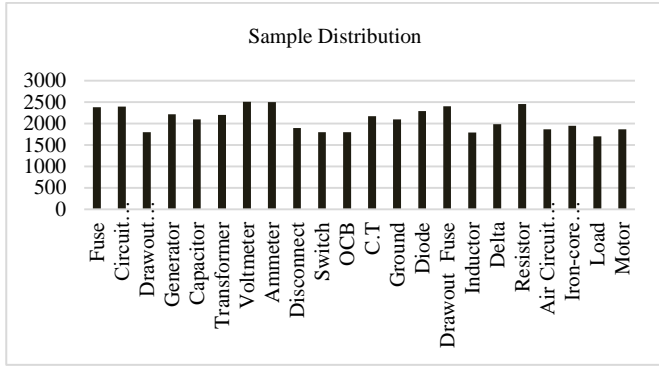


Fig. 12. Sample distribution in MFC-GAN generated dataset.

In this study, MFC-GAN models were able to produce occurrences of minority classes that were significantly underrepresented in the dataset, as demonstrated in Fig. 12. Both subjective and objective methods were used to evaluate the generated samples, including an assessment of the classifier's performance before and after incorporating the samples into the training sets. The results showed an improvement in performance across several commonly used evaluation parameters. However, it should be acknowledged that the class imbalance issue can only be addressed to some extent by MFC-GAN, and other strategies may need to be explored and utilized.

The study categorizes the outcome of classification predictions into four groups based on the relationship between the predicted output and the actual value: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). To assess the efficacy of defect detection, the study calculates the precision, recall, and F1 score of the model for different types of faults. Precision rate, which measures the accuracy of detection findings, is computed by dividing the number of symbols expected to be positive by the total number of symbols predicted to be positive. Recall rate, which gauges the thoroughness of detection findings, is calculated by dividing the number of samples expected to be positive by the total number of samples that actually have a positive value. Precision and recall are given in Equations (14) and (15):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (15)$$

To evaluate classification problems, it is essential to take into account both the accuracy of identification and the completeness of detection. The model is evaluated using the F1 score, which considers both precision and recall. Equations (16) and (17) express accuracy as the number of symbols that are correctly identified.

$$\text{F1-score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

$$\text{Accuracy} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

The results of the detection data are shown in Table II. The attributes represent the actual class, and each row represents the predicted category. The total number of symbols for each category is calculated by adding up the numbers in each column. The predicted category and the total number of predicted symbols for that category are shown in each row. Our proposed method can accurately detect the majority of single line-based engineering symbols. In this study, the precision of symbol detection for all types is over 95%, the average recall rate is 93.67%, and the F1 score is above 0.9. The average frame detection time is 0.074 seconds, while the average recall and precision rates are 90.67% and 0.074 seconds, respectively.

TABLE II. SYMBOLS GENERATED USING MFC-GAN

Symbol Name	Symbol	Original Instances	Generated Instances
Fuse		480	2380
Circuit Breaker		344	2400
Drawout Circuit Breaker		790	1800
Capacitor		821	2011
Generator		1340	2219
Transformer		355	2200
Voltmeter		212	2587
Ammeter		740	2500
Disconnect		99	1900
Switch		616	1800
OCB		799	1800
C.T		389	2178
Ground		655	2100
Diode		340	2291
Drawout Fuse		589	2408
Inductor		203	1790
Delta		840	1990
Resistor		700	2455
Air Circuit Breaker		780	1870
Iron-core Inductor		750	1950
Load		117	1701
Motor		541	1870

V. DISCUSSION

This section may be divided into two subsections which provide comparisons and conclusive remarks on the experiments.

A. Comparison with Different Data Augmentation Techniques

Table II displays the frequency of the different classes in both datasets. After generating synthetic SLD images using MFC-GAN, it can be observed that the new samples are approximately balanced. However, to address the issue of class imbalance, the original dataset could be improved by including additional distinct and separate photos.

TABLE III. YOLOV5 CLASSIFICATION PERFORMANCE OF SYMBOLS ON AUGMENTED DATASET

Metric	Switch	Relay	Motor	Generator	Load	Inductor	Fuse	Resistor
Precision	1.00	1.00	0.85	0.89	1.00	1.00	1.00	1.00
Recall	0.94	0.89	0.92	1.00	1.00	0.90	0.91	0.87
F1-score	0.89	0.94	0.91	0.95	0.88	0.90	0.92	0.91
Accuracy	1.00	0.98	0.99	1.00	1.00	1.00	1.00	1.00

We conducted an experiment to test the accuracy and performance of YOLOv8 in various conditions and configurations using 800 images, and an example of the detection results can be seen in Fig. 13. The accuracy testing and performance of the experiment using images from our datasets are displayed in Table IV. YOLOv5 is generally more accurate than its recent version. Group 2, which is the augmented dataset, had the highest average accuracy of 96% when using YOLOv5, with only five detection errors. YOLO's performance can be improved by utilizing a large dataset that includes both real and synthetic images generated by GANs. When a deep learning-based method is trained on a small and insufficient dataset, it may result in overfitting and difficulties in mapping the object [52]. Adding noise or generating fake images during training can make the process of learning the input image from the output image easier, reducing general errors, and enhancing the training component [53]. Thus, to improve item identification accuracy, it is necessary to include synthetic images in addition to actual photographs.

B. Missed Detection

Table V presents the outcomes of the evaluation of the model on the SLD components dataset, revealing that there were a total of 52 instances where the SLD components were either absent or wrongly classified as some other symbols. Out of these occurrences, eight symbols were misclassified, while the remaining 46 symbols were not detected entirely. This issue can be attributed in part to the nature of some drawings, where symbols are nearly completely obscured by text and comments. The Intersection over Union (IOU) metric in Table V confirms that these missed symbols have a zero IOU, indicating that they were not detected by the model.

TABLE IV. COMPARISON OF YOLOV8 AND YOLOV5 CLASS ACCURACY

Dataset	Accuracy	Wrong Detection	Missed Detection
Original	95%	8	46
Augmented	96.3%	3	2

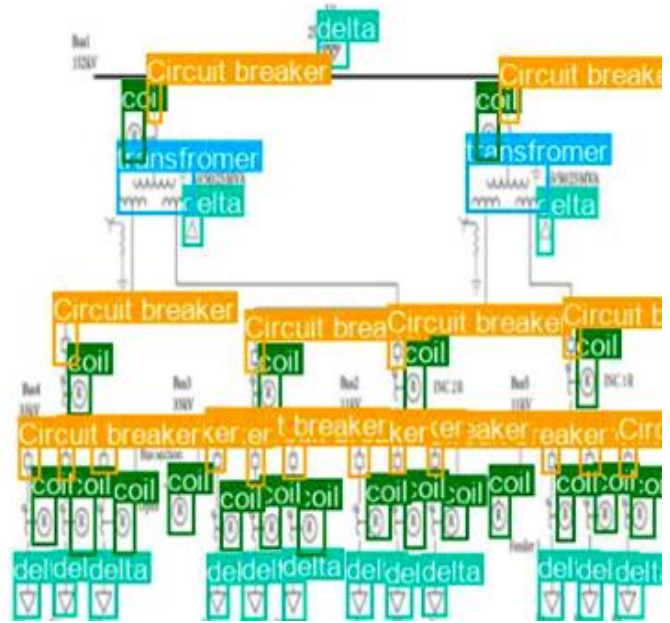


Fig. 13. Symbols detected in augmented dataset.

After conducting a visual inspection of the data in Table V, some symbols were found to be mislabeled. Specifically, when reviewing the results for the switch representation, it was observed that the algorithm had predicted the correct class for symbols with an incorrect label. However, for some symbols, the algorithm had predicted the wrong class label.

TABLE V. SYMBOLS MISSED OR OVERLOOKED BY THE CLASSIFIERS

Class	No of Sample Occurrence		Predicted Sample Class		IOU	
	YOLOv8	YOLOv5	YOLOv8	YOLOv5	YOLOv8	YOLOv5
Disconnect	2	5	Switch	Ground	0.50	0.91
Load	1	3	-	Fuse	0.56	1.00
Inductor	2	4	-	-	-	0.00
Voltmeter	-	16	-	-	-	0.00
Diode	-	10	-	-	-	0.00
Circuit Breaker	-	8	-	-	-	0.00

VI. CONCLUSION AND FUTURE RECOMMENDATION

In this study, we proposed a system for analyzing and processing complex engineering drawings. Our approach achieved more than 96% accuracy in recognizing symbols on the drawings, based on extensive testing on a large collection of SLD sheets provided by an industry partner. We utilized advanced bounding-box detection techniques, which

demonstrated high accuracy in identifying symbols from 22 different categories, despite some of these symbols having only minor differences. To address class imbalance in the symbol dataset, we suggested a GAN-based model. Our experiments showed that our system could generate realistic engineering symbols, and adding this synthetic data to the training set improved classification accuracy. According to the experimental results, the proposed GAN model was capable of learning from a smaller amount of training in-stances.

The subsequent emphasis of this study will be on the application of GANs to the creation of symbols in a schematic environment. In future development, an integrated system will be created using the recommended methods to enable thorough analysis and processing of technical diagrams like SLD. This approach will make it much easier to perform additional tasks such as line detection or text localization. Furthermore, future work will involve combining Explainable AI (XAI) and other GAN methods such as WGAN, CycleGAN, PCCGAN, and StyleGAN with other detection techniques.

ACKNOWLEDGMENT

We appreciate Yayasan UTP FRG (YUTP-FRG), grant number 015LC0-280, and Computer and Information Science Department of Universiti Teknologi PETRONAS for providing funding and support for this research.

REFERENCES

- [1] BHANBHRO, H., HOOL, Y. K., HASSAN, Z., & SOHU, N., "Modern Approaches towards Object Detection of Complex Engineering Drawings," in Proc. International Conference on Digital Transformation and Intelligence (ICDI), IEEE, 2022.
- [2] E. ELYAN, L. JAMIESON, AND A. ALI-GOMBE, "Deep learning for symbols detection and classification in engineering drawings," Neural networks, vol. 129, pp. 91-102, 2020.
- [3] X. Y. , G. F. MENG, AND C. H. PAN, "Scene text detection and recognition with advances in deep learning: a survey," (in English), Int J Doc Anal Recog, vol. 22, no. 2, pp. 143-162, Jun 2019, doi: 10.1007/s10032-019-00320-5.
- [4] S. MANI, M. A. HADDAD, D. CONSTANTINI, W. DOUHARD, Q. W. LI, and L. Poirier, "Automatic Digitization of Engineering Diagrams using Deep Learning and Graph Search," (in English), Ieee Comput Soc Conf, pp. 673-679, 2020, doi: 10.1109/Cvprw50498.2020.00096.
- [5] C. F. MORENO-GARCIA, E. ELYAN, AND C. JAYNE, "New trends on digitisation of complex engineering drawings," (in English), Neural Comput Appl, vol. 31, no. 6, pp. 1695-1712, Jun 2019, doi: 10.1007/s00521-018-3583-1.
- [6] T. M. NGUYEN, L. V. PHAM, C. C. NGUYEN, AND V. V. NGUYEN, "Object Detection and Text Recognition in Large-scale Technical Drawings," (in English), Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (Icpram), pp. 612-619, 2021, doi: 10.5220/0010314406120619.
- [7] J. K. NURMINEN, K. RAINIO, J.-P. NUMMINEN, T. SYRJÄNEN, N. PAGANUS, AND K. HONKOILA, "Object detection in design diagrams with machine learning," in International Conference on Computer Recognition Systems, 2019: Springer, pp. 27-36.
- [8] A. REZVANIFAR, M. COTE, AND A. B. ALBU, "Symbol Spotting on Digital Architectural Floor Plans Using a Deep Learning-based Framework," (in English), Ieee Comput Soc Conf, pp. 2419-2428, 2020, doi: 10.1109/Cvprw50498.2020.00292.
- [9] S. SARKAR, P. PANDEY, AND S. KAR, "Automatic Detection and Classification of Symbols in Engineering Drawings," arXiv preprint arXiv: 2204.13277, 2022.
- [10] Q. S. Wang, F. S. Wang, J. G. Chen, and F. R. Liu, "Faster R-CNN Target-Detection Algorithm Fused with Adaptive Attention Mechanism," (in Chinese), Laser Optoelectron P, vol. 59, no. 12, Jun 2022, doi: 10.3788/Lop202259.1215016.
- [11] L. H. WEN AND K. H. JO, "Fast LiDAR R-CNN: Residual Relation-Aware Region Proposal Networks for Multiclass 3-D Object Detection," (in English), Ieee Sens J, vol. 22, no. 12, pp. 12323-12331, Jun 15 2022, doi: 10.1109/Jsen.2022.3172446.
- [12] CINTRA, R. J., DUFFNER, S., GARCIA, C., AND LEITE, AL., "Low-complexity Approximate Convolutional Neural Networks," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 12, pp. 5981-5992, 2018.
- [13] KHAN, S. H., HAYAT, M., BENNAMOUN, M., SOHEL, F. A., AND TOGNERI, R., "Cost-sensitive Learning of Deep Feature Representations from Imbalanced Data," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 8, pp. 3573-3587, Aug. 2018.
- [14] STUHLSATZ, A., LIPPEL, J., & ZIELKE, "Feature Extraction with Deep Neural Networks by a Generalized Discriminant Analysis," IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 4, pp. 596-608, Apr. 2012.
- [15] REN, S., HE, K., GIRSHICK, R., & SUN, J., "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks," in Proc. NIPS, 2015, pp. 91-99.
- [16] REDMON, J., DIVVALA, S., GIRSHICK, R., & FARHADI, A., "You Only Look Once: Unified, Real-time Object Detection," in Proc. CVPR, 2016, pp. 779-788.
- [17] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proc. CVPR, 2005, pp. 886-893.
- [18] P. F. FELZENSZWALB et al., "Object Detection with discriminatively Trained Part-based Models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627-1645, Sep. 2010.
- [19] M. EVERINGHAM et al., "The Pascal Visual Object Classes (VOC) Challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303-338, 2008.
- [20] G. E. HINTON AND R. R. SALAKHUTDINOV, "Reducing the Dimensionality of Data with Neural Networks," Science, vol. 313, no. 5786, pp. 504507, 2006.
- [21] VIG, E., DORR, M., & COX, D., "Large-scale Optimization of Hierarchical Features for Saliency Prediction in Natural Images," in Proc. CVPR, 2014, pp. 2798-2805.
- [22] BHANBHRO, H., HOOL, Y. K., HASSAN, Z., & SOHU, N., "Modern Deep Learning Approaches for Symbol Detection in Complex Engineering Drawings," in Proc. International Conference on Digital Transformation and Intelligence (ICDI), IEEE, 2022.
- [23] ELYAN, E., MORENO-GARCÍA, C. F., & JOHNSTON, P., "Symbols in Engineering Drawings (SIED): An Imbalanced Dataset benchmarked by Convolutional Neural Networks," in Proc. 21st EANN (Engineering Applications of Neural Networks), 2020.
- [24] RICA, E., ALVAREZ, S., MORENO-GARCIA, C. F., & SERRATOSA, F., "Zero-Error Digitisation and Contextualisation of Piping and Instrumentation Diagrams Using Node Classification and Sub-graph Search," Springer International Publishing, August 26-27, 2023.2.6
- [25] GUPTA, M., WEI, C., & CZERNIAWSKI, T., "Automated Valve Detection in Piping and Instrumentation (P&ID) Diagrams," in Proc. International Symposium on Automation and Robotics in Construction. Vol. 39. IAARC Publications, 2022.
- [26] SHEN, C., LV, P., MAO, M., LI, W., ZHAO, K., & YAN, Z., "Substation One-Line Diagram Automatic Generation Based On Image Recognition," in Proc. Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT). IEEE, 2022.
- [27] A. ALI-GOMBE AND E. ELYAN, "MFC-GAN: Class-imbalanced dataset classification using Multiple Fake Class Generative Adversarial Network," (in English), Neurocomputing, vol. 361, pp. 212-221, Oct 7 2019, doi: 10.1016/j.neucom.2019.06.043.
- [28] E. ELYAN, L. JAMIESON, AND A. ALI-GOMBE, "Deep learning for symbols detection and classification in engineering drawings," (in English), Neural Networks, vol. 129, pp. 91-102, Sep 2020, doi: 10.1016/j.neunet.2020.05.025.
- [29] V. NAOSEKIPAM AND N. SAHU, "Text detection, recognition, and script identification in natural scene images: a Review," (in English), Int

- J Multimed Inf R, vol. 11, no. 3, pp. 291-314, Sep 2022, doi: 10.1007/s13735-022-00243-8.
- [30] H. BHANBHRO, S. R. HASSAN, S. Z. NIZAMANI, S. T. BAKHS, AND M. O. ALASSAFI, "Enhanced Textual Password Scheme for Better Security and Memorability," (in English), *Int J Adv Comput Sc*, vol. 9, no. 7, pp. 209-215, Jul 2018.
- [31] R. HUANG, J. GU, X. SUN, Y. HOU, AND S. UDDIN, "A rapid recognition method for electronic components based on the improved YOLO-V3 network," *Electronics*, vol. 8, no. 8, p. 825, 2019.
- [32] H. LEE, J. LEE, H. KIM, AND D. MUN, "Dataset and method for deep learning-based reconstruction of 3D CAD models containing machining features for mechanical parts," *Journal of Computational Design and Engineering*, vol. 9, no. 1, pp. 114-127, 2022.
- [33] S. E. WHANG, Y. ROH, H. SONG, AND J.-G. LEE, "Data collection and quality challenges in deep learning: A data-centric ai perspective," *The VLDB Journal*, pp. 1-23, 2023.
- [34] J. WANG, Y. CHEN, Z. DONG, AND M. GAO, "Improved YOLOv5 network for real-time multi-scale traffic sign detection," *Neural Computing and Applications*, pp. 1-13, 2022.
- [35] C. F. MORENO-GARCÍA, E. ELYAN, AND C. JAYNE, "Heuristics-based detection to improve text/graphics segmentation in complex engineering drawings," in *Engineering Applications of Neural Networks: 18th International Conference, EANN 2017, Athens, Greece, August 25–27, 2017, Proceedings, 2017: Springer*, pp. 87-98.
- [36] L. JAMIESON, C. F. MORENO-GARCIA, AND E. ELYAN, "Deep learning for text detection and recognition in complex engineering diagrams," in *2020 International Joint Conference on Neural Networks (IJCNN), 2020: IEEE*, pp. 1-7.
- [37] M. F. THEISEN, K. N. FLORES, L. S. BALHORN, AND A. M. SCHWEIDTMANN, "Digitization of chemical process flow diagrams using deep convolutional neural networks," *Digital Chemical Engineering*, vol. 6, p. 100072, 2023.
- [38] M. KARTHI, V. MUTHULAKSHMI, R. PRISCILLA, P. PRAVEEN, AND K. VANISRI, "Evolution of yolo-v5 algorithm for object detection: automated detection of library books and performance validation of dataset," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), 2021: IEEE*, pp. 1-6.
- [39] Naosekham, V.; Sahu, N. Text detection, recognition, and script identification in natural scene images: a Review. *Int J Multimed Inf R* 2022, 11 (3), 291-314. DOI: 10.1007/s13735-022-00243-8.
- [40] Zhang, Q. R.; Zhang, M.; Chen, T. H.; Sun, Z. F.; Ma, Y. Z.; Yu, B. Recent advances in convolutional neural network acceleration. *Neurocomputing* 2019, 323, 37-51. DOI: 10.1016/j.neucom.2018.09.038.
- [41] Antoniou, A.; Storkey, A.; Edwards, H. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340* 2017.
- [42] Amur, Z. H.; Hooi, Y.; Sodhar, I. N.; Bhanbhro, H.; Dahri, K. State-of-the Art: Short Text Semantic Similarity (STSS) Techniques in Question Answering Systems (QAS). In *International Conference on Artificial Intelligence for Smart Community: AISC 2020, 17–18 December, Universiti Teknologi Petronas, Malaysia, 2022; Springer*: pp 1033-1044.
- [43] Amur, Z. H.; Kwang Hooi, Y.; Bhanbhro, H.; Dahri, K.; Soomro, G. M. Short-Text Semantic Similarity (STSS): Techniques, Challenges and Future Perspectives. *Applied Sciences* 2023, 13 (6), 3911.
- [44] Baur, C.; Albarqouni, S.; Navab, N. MelanoGANs: high resolution skin lesion synthesis with GANs. *arXiv preprint arXiv:1804.04338* 2018.
- [45] Buda, M.; Maki, A.; Mazurowski, M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks* 2018, 106, 249-259.
- [46] Denton, E. L.; Chintala, S.; Fergus, R. Deep generative image models using a² laplacian pyramid of adversarial networks. *Advances in neural information processing systems* 2015, 28.
- [47] Dong, Q.; Gong, S.; Zhu, X. Class rectification hard mining for imbalanced deep learning. In *Proceedings of the IEEE international conference on computer vision, 2017*; pp 1851-1860.
- [48] Dosovitskiy, A.; Springenberg, J. T.; Riedmiller, M.; Brox, T. Discriminative unsupervised feature learning with convolutional neural networks. *Advances in neural information processing systems* 2014, 27.
- [49] Douzas, G.; Bacao, F. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Syst Appl* 2018, 91, 464-471.
- [50] Fernández, A.; López, V.; Galar, M.; Del Jesus, M. J.; Herrera, F. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowledge-based systems* 2013, 42, 97-110.
- [51] Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. Synthetic data augmentation using GAN for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), 2018; IEEE*: pp 289-293.
- [52] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition, 2016*; pp 770-778.
- [53] Huang, C.; Li, Y.; Loy, C. C.; Tang, X. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition, 2016*; pp 5375-5384.
- [54] Inoue, H. Data augmentation by pairing samples for images classification. *arXiv preprint arXiv:1801.02929* 2018.
- [55] Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* 2017.
- [56] Mariani, G.; Scheidegger, F.; Istrate, R.; Bekas, C.; Malossi, C. Bagan: Data augmentation with balancing gan. *arXiv preprint arXiv:1803.09655* 2018.
- [57] Netzer, Y.; Wang, T.; Coates, A.; Bissacco, A.; Wu, B.; Ng, A. Y. Reading digits in natural images with unsupervised feature learning. 2011.
- [58] Odena, A. Semi-supervised learning with generative adversarial networks. *arXiv preprint arXiv:1606.01583* 2016.
- [59] Wan, L.; Wan, J.; Jin, Y.; Tan, Z.; Li, S. Z. Fine-grained multi-attribute adversarial learning for face generation of age, gender and ethnicity. In *2018 International Conference on Biometrics (ICB), 2018; IEEE*: pp 98-103.
- [60] Yue, Y.; Liu, H.; Meng, X.; Li, Y.; Du, Y. Generation of high-precision ground penetrating radar images using improved least square generative adversarial networks. *Remote Sensing* 2021, 13 (22), 4590.
- [61] Thuan, D. Evolution of Yolo algorithm and Yolov5: The State-of-the-Art object detection algorithm. 2021.

The Spatial Distribution of Atmospheric Water Vapor Based on Analytic Hierarchy Process and Genetic Algorithm

Fengjun Wei¹, Chunhua Liu², Rendong Guo^{3*}, Xin Li⁴, Jilei Hu⁵, Chuanxun Che⁶

School of Energy and Building Engineering, Shandong Huayu University of Technology, Dezhou, 253034, China^{1,2,3,6}
Research and Development Center of Building Energy Saving Engineering Technology, Shandong Huayu University of Technology, Dezhou, 253034, China^{1,2,3}

School of Civil Engineering, Shenyang University, Shenyang, 110044, China⁴

Key Laboratory of Geological Hazards on Three Gorges Reservoir Area-Ministry of Education, China Three Gorges University, Yichang, 443002, China⁵

Abstract—The inversion of water vapor spatial distribution using ground-based global navigation satellite systems is a technique that utilizes the propagation delay of satellite signals in the atmosphere to retrieve atmospheric water vapor information. To further promote the accuracy of the information obtained by this method, a satellite system is designed to solve the spatial distribution of atmospheric water vapor based on chromatography technology and genetic algorithm. Firstly, the accuracy of the empirical air temperature and pressure model to calculate the zenith statics delay is analyzed. To optimize the global weighted average temperature model, a model that considers the decreasing rate of atmospheric weighted average temperature and a model based on the linear relationship between surface heat and weighted average temperature are proposed. The idea of removal interpolation restoration is introduced to achieve regional interpolation of atmospheric precipitable water. Finally, in response to the problem of multiple solutions in the current water vapor chromatography equation, a genetic algorithm based chromatography method is put forward to achieve the solution of atmospheric water vapor spatial distribution. The experimental analysis shows that the average root mean square error and average absolute error of the design method of the research institute are 1.78g/m³ and 1.41g/m³, respectively, which can realize the calculation of atmospheric water vapor density distribution with high accuracy.

Keywords—Global navigation satellite system; spatial distribution of water vapor; genetic algorithm; chromatography technology

I. INTRODUCTION

The atmosphere is a crucial carrier of the earth's gas and water cycle, providing oxygen for humans and animals, carbon dioxide for plant photosynthesis, and water for all life [1]. Through the greenhouse effect, the atmosphere maintains a suitable temperature on the Earth, and through the absorption of harmful rays by ozone, protects earth's organisms from harm. Water vapor is an important component of the atmosphere, mainly concentrated in the lower atmosphere. It has an important impact on the atmospheric greenhouse effect and circulation [2,3]. The atmosphere water vapor has complex temporal and spatial changes, and accurately measuring, modeling, and predicting the content

and changes of water vapor is of great importance for studying the greenhouse effect and climate change. With the continuous expansion of Global Navigation Satellite System (GNSS) and its related fields, GNSS technology used to study the atmosphere water vapor spatial distribution (AWVSD) has made significant progress [4,5]. To achieve a more accurate solution of water vapor distribution, research is conducted on the estimation of atmospheric precipitable water volume (PWV) based on GNSS technology, and the idea of removal interpolation restoration is introduced to optimize the accuracy of the calculation. In terms of three-dimensional water vapor chromatography, a chromatography manner with genetic algorithm (GA) is proposed to address the problems in the chromatography equation.

The gap between research and existing national and international research: Currently, there are generally four main methods for calculating atmospheric water vapor content. The first method is direct calculation using measured sounding data. Sounding data are relatively old, objective and accurate, so they are often used to verify other calculation methods. However, due to the limited number of sounding stations, it is not possible to describe the spatiotemporal distribution of regional water vapor in detail, especially for areas with uneven terrain loads and stations. The second is to use remote sensing data to invert the atmospheric water vapor content, a new technology that has been developed and widely used in recent years, but whose measurement accuracy still needs to be improved. The third is to establish the relationship between the water vapor content and the surface meteorological elements for the calculation. The calculation is simple and the results are ideal, but this greatly increases the number of stations. Fourth, NCEP/NCAR reanalysis data is currently the most widely used method for calculating water vapor content. Gridded data can compensate for the shortcomings of insufficient stations in calculating the spatial distribution of water vapor, but the coarse grid affects the fine analysis and its applicability is not high enough. The method developed by the Research Institute is an improved method based on the use of remote sensing data to invert the atmospheric water vapor content, ensuring high accuracy in the spatial distribution of water vapor with fewer stations.

Motivation and potential benefits of the research: Over the past century, the global climate has undergone a significant change characterised by global warming, and the warming of the climate system has become an indisputable fact. And global warming has also led to more frequent natural disasters caused by extreme weather events, with extremely high losses for society. Extreme precipitation often leads to more severe flood disasters. Detecting water vapor in real time and controlling its spatiotemporal changes, as well as studying its interaction with precipitation, is of great research importance for precipitation forecasting. However, traditional water vapor detection methods have drawbacks such as high operational costs, large station requirements and insufficient accuracy. The use of GNSS to invert the spatial distribution of water vapor has good performance, but the accuracy needs to be improved. Research is underway to improve this technology to achieve high performance water vapor spatial distribution detection.

The main contributions and impacts of the research include:

1) Using different model expressions, different practical resolutions, and different spatial resolution modeling data, the impact characteristics of different modeling factors on the accuracy of empirical temperature and pressure models were studied. A temperature and pressure model based on segmented practical ideas was constructed, and the accuracy of calculating ZHD estimates for different ZHDs in measured meteorological parameters was evaluated within the context of de Qianqiu.

2) The weighted average temperature decline rates for different height intervals were obtained and analyzed, and a CPT2wh model was established. At 1 × Provide the average value, annual amplitude, and semi-annual amplitude of Tm decline rate on a grid point. The proposed model effectively improves the lack of elevation correction in Tm estimation and significantly improves the accuracy of Tm estimation for stations with large elevation differences.

3) A new method for detecting the spatial distribution of water vapor based on genetic algorithm and Analytic Hierarchy Process has been proposed. This method does not need to rely too much on the conditions of the Moon Lake, prior information and external meteorological data, and converts the inverse problem of the matrix into the function optimization problem, which effectively solves the ill conditioned problem of the equation.

4) A GNSS-PWV spatial interpolation method based on the idea of removal interpolation recovery is proposed. This method does not require the surface measured meteorological data of the station to be measured, nor does it require regression analysis of the observed data, which can effectively avoid the issue of elevation correction in PWV interpolation.

The research and work can address the shortcomings of traditional methods of spatial detection of water vapour and provide more accurate detection results. The data will be useful as a guide for government planning and regional responses to climate change.

The specific content of the study is divided into three parts:

The first part is the literature review section, which analyzes the current research status at home and abroad, and explores the key points that need to be overcome in the research.

The second part is the methodology section, which mainly introduces the technologies required for the AWVSD calculation method designed by the research institute, and makes improvements to address the limitations of related technologies.

The third part is the experimental analysis section, which mainly analyzes the performance of the research institute's design methods.

II. RELATED WORKS

The amount of water content in clouds is an important parameter for studying the impact of clouds on climate. Many scholars have used different methods to calculate and analyze the AWVSD characteristics. Shi et al. used the HYSPLIT platform to simulate the Lagrangian trajectory of air envelopment in East China during the summer monsoon. It investigated four different periods during its seasonal migration from south to north [6]. Huang et al. used radio temperature measurement data from 2012 to 2017 to establish an empirical model of atmospheric weighted average temperature (Tm) in Guilin, China. Then, they used observation data from 11 GNSS stations in Guilin to study the spatiotemporal characteristics of GNSS derived PWV under heavy rain from June to July 2017 [7]. Lee et al. found in their observation of the time process of H2O2 generation in condensate droplets that it was typically generated from droplets smaller than 10 microns [8]. Tan et al. obtained the land surface temperature (LST) of Dongting Lake (China) from Landsat 7 data, and discussed its relationship with land cover (LULC) type. The outcomes denoted that LST varied greatly among different LULC types, with higher LST in building areas and lower LST in other areas. Water bodies played a critical supervision role in reducing LST [9].

GNSS technology was gradually developing and becoming a focus of research for scholars. Pan et al. considered the advantages and disadvantages of remote sensing technology and Global Navigation Satellite System Interferometric Reflection (GNSS-IR) in trajectory recognition of water content (VWC), and proposed a point surface fusion method with GA combined with backpropagation neural network (GA-BP) for GNSS IR and MODIS data to raise the accuracy of VWC estimation [10]. Zheng et al. used the signal-to-noise ratio of navigation satellite signals to invert sea level based on observation data from MAYG on the east coast of Africa from 2017 to 2019 [11]. LÜ et al. used Differential Interferometric Synthetic Aperture Radar (InSAR) and pixel offset tracking to gain the line of sight displacement and near-field range displacement of the M6.9 earthquake in Menyuan, Qinghai from SAR images. At the same time, they obtained high-speed displacement waveforms of 16 GNSS stations through historically accurate point positioning schemes and inverted the seismic fracture process [12]. Lewen et al. conducted

monitoring of changes in the spatial environment of the line of sight based on global and regional GNSS reference station networks for inversion of tropospheric and ionospheric parameters [13].

GA is an efficient, parallel, and global search method. Jalali Z et al. used GAs and constant development manners to optimize the exterior walls of an office building. The SPEA-2 algorithm was utilized for multi-objective optimization [14]. Garud et al. extensively reviewed the applicability of artificial neural networks (ANN), fuzzy logic (FL), and GA, as well as their hybrid models, using AI for effectiveness prediction. Besides, some literature on predicting solar radiation using ANN, FL, GA, and their hybrid models has been summarized [15]. Shariati et al. collected a database of 1030 datasets to intelligently predict the strength of concrete containing slag and fly ash as partial substitutes for cement [16]. Abualigah et al. proposed an efficient task scheduling optimization method based on a multiverse optimizer and genetic algorithm (MVO-GA). MVO-GA was utilized to improve the effect of task transmission through cloud networks, provide appropriate transfer decisions, and rearrange transfer tasks with the efficiency weights collected in the cloud [17].

Based on the above literature analysis, there are various methods for calculating and analyzing the AWVSD, but the accuracy of the obtained AWVSD analysis methods still needs to be improved. To this end, the study utilizes GNSS technology to estimate atmospheric PWV, and then uses GA algorithm for water vapor chromatography. A solution model for AWVSD based on Analytic Hierarchy Process (AHP) and GA is obtained.

III. A SOLUTION MODEL FOR AWVSD BASED ON AHP AND GA

Water vapor can effectively absorb solar longwave radiation and surface infrared radiation, thereby providing insulation for the entire Earth system. Its content in the atmosphere has complex spatial and temporal variations, constantly affecting weather characteristics and climate environments around the world. Therefore, the accurate measurement, modeling and prediction of water vapor levels and changes are of great importance for the study of the greenhouse effect and climate change. At the same time, an in-depth understanding of the spatiotemporal changes of water vapor plays an important role in improving the accuracy of weather forecasting and conducting disaster weather warnings. When electromagnetic waves pass through the Atmosphere of Earth, they are affected by the ionospheric delay and the flow delay, which will lead to the extension of the electromagnetic wave propagation practice and the bending of the propagation path. Therefore, monitoring and studying the distribution of water vapor in the atmosphere is of great importance for services such as radio communications, navigation, positioning and timing. In order to improve the accuracy of information obtained from current water vapor spatial distribution inversion techniques, research has been carried out on key technologies for obtaining atmospheric precipitable water and tomographic inversion of atmospheric water vapor density, with a focus on ground-based GNSS inversion. Firstly, the empirical air temperature and pressure model is analyzed

to calculate the accuracy of the zenith Statics delay. In order to optimize the global weighted average temperature model, the model of worrying about the decline rate of the atmospheric weighted average temperature and the model based on the linear relationship between the surface heat and the weighted average temperature are proposed. And introduce the idea of removal interpolation restoration to achieve regional interpolation of atmospheric precipitable water. Finally, in response to the problem of multiple solutions in the current water vapor chromatography equation, a genetic algorithm based tomography method is proposed to achieve the solution of atmospheric water vapor spatial distribution.

1) *GNSS-PWV spatial interpolation based on the idea of removal interpolation recovery*: In the GNSS solution model, each satellite signal corresponds to a tropospheric delay, and directly estimating them will result in rank deficiency in the equation. The convective delay is modeled as the sum of zenith directional delay (ZTD) and the product of atmospheric horizontal gradient and their corresponding projection functions. The calculation method of ZTD is shown in equation (1).

$$STD = m_h(ele) \times m_w(ele) \times ZWD + m_\Delta(ele) \times \cot(ele) \times [G_{NS} \cos(azi) + G_{WE} \sin(azi)] \quad (1)$$

In formula (1), *STD* means the total delay of tropospheric oblique path; m_h stands for the dry mapping function; m_w stands for the wet mapping function; m_Δ expresses the atmospheric horizontal gradient mapping function; G_{NS} and G_{WE} express the atmospheric horizontal gradient in the north-south and east-west directions respectively; *ele* and *azi* indicate the satellite altitude angle and azimuth angle respectively; *ZWD* denotes the zenith wet delay; *ZHD* means the zenith statics delay. Usually, atmospheric precipitable water vapor (PWV) and integrated water vapor (IWV) are utilized to characterize the information of atmospheric water vapor content. The conversion relationship between the two is shown in equation (2).

$$IWV = \rho_w \times PWV \quad (2)$$

In equation (2), ρ_w means the density of liquid water. ZWD and PWV's conversion relationship is shown in equation (3).

$$\left\{ \begin{array}{l} PWV = \Pi \cdot ZWD \\ \Pi = \frac{10^6}{\rho_w \times R / m_w \times (k_2' + k_3 / T_m)} \end{array} \right. \quad (3)$$

In equation (3), Π means the conversion factor; R denotes the general gas constant; m_w indicates the molar mass of wet air; k_2' and k_3 express the atmospheric refractive index constants; T_m means the atmospheric weighted average temperature. The process of inverting PWV through ground GNSS is shown in Fig. 1.

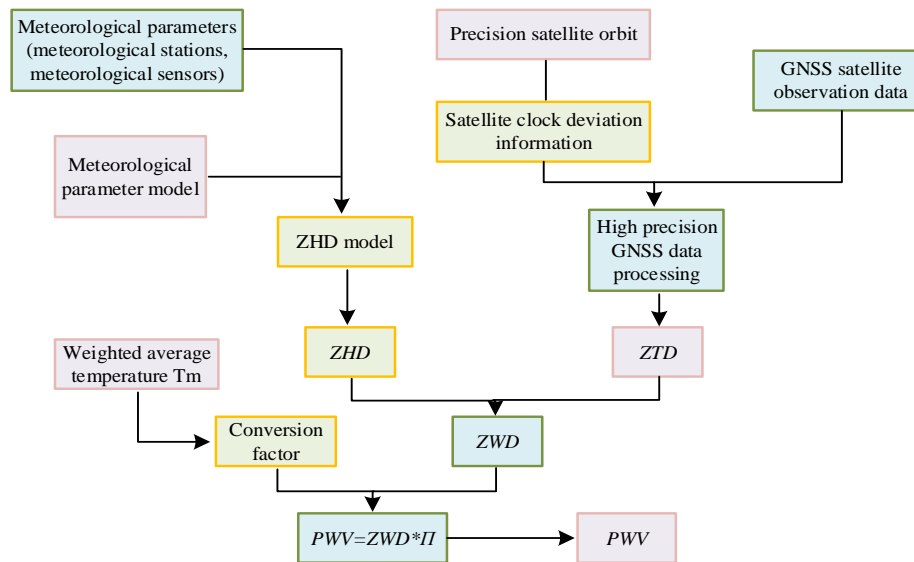


Fig. 1. Process of PWV inversion based on ground GNSS.

After obtaining the GNSS station ZWD, it is necessary to use conversion factors to obtain the atmospheric PWV. It is a function of the weighted T_m , and accurately obtaining T_m is the useful way to improving the accuracy of GNSS-PWV. To this end, the GPT2w model is applied to calculate T_m . When using GPT2w to calculate T_m , the study first selects the four model grid points closest to the desired station, and uses the model coefficients stored in external grid files to calculate the T_m estimates of the four grid points. Then, the interpolation algorithm is used to interpolate the T_m estimates of the four grid points to gain the T_m estimates of the station to be measured. However, during the research, it is found that the GPT2w model lacks elevation correction when calculating T_m , which limits the accuracy of the model in some cases, especially on the waiting points with significant elevation differences from the GPT2w grid points. For this purpose, the study adopts the index of virtual temperature, temperature decreasing rate, and water vapor pressure (WVP) decreasing factor to adjust the pressure, temperature, and WVP accordingly. At the same time, it applies the vertical T_m decline rate to the GPT2w model. Considering the limitations of the relationship between T_m and surface temperature (T_s) used globally, a GGTm T_s model is established using GGOSAtmosphere and ECMWF data. By providing high-precision T_m - T_s relationships on global grid points, more accurate T_m estimation can be achieved.

GNSS technology is used to gain PWV, which has become an important data source for meteorological departments. However, due to environmental and economic factors, GNSS stations are often scattered and have large distances, which limits the analysis of AWVSD in certain applications. Therefore, research is conducted on spatial interpolation of GNSS-PWV. It has a closed connection between water vapor and terrain, so it needs to consider the impact of terrain factors on PWV interpolation. To address the above issues, a method based on the idea of removal interpolation restoration is

proposed for spatial interpolation of GNSS-PWV. The interpolation method is shown in Fig. 2.

Using the coordinates and time information of the measuring station, the GPT2w model can calculate meteorological parameters, and use these two parameters to estimate the ZWD of the measuring station. The calculation method is shown in equation (4).

$$ZWD = 10^{-6} \times \left(k_2' \times k_3 / T_m \right) \times \frac{R / m_d \times g_m}{(\lambda + 1)} \times e \quad (4)$$

In equation (4), m_d means the molar mass of dry air; g_m indicates the average gravity; e and λ denote the WVP and the decline rate of WVP respectively. After obtaining ZWD, it is converted into PWV using a conversion factor. The conversion method is shown in equation (5).

$$PWV = \frac{m_w \times g_m}{m_d \times p_w \times (\lambda + 1)} \times e \quad (5)$$

Using the GPT2w model, the PWV estimation at the corresponding time of the station calculated by using the above formulas (4) and (5) is denoted as GPT2w_PWV. The PWV true value obtained by using GNSS technology to high-precision process the observation data is denoted as GNSS_PWV. Firstly, it calculates the difference between the two types of PWVs based on PWV_residual. Then, the interpolation algorithm is used to interpolate the PWV_residual of the GNSS station to obtain the difference of the desired station. Finally, the obtained difference and its GPT2w_PWV are interpolated into the PWV of the station to be tested. This method achieves interpolation without the need for surface meteorological observation data and regression fitting processes, and considers the influence of elevation issues on PWV interpolation. The distribution and elevation of GNSS stations are shown in Fig. 3.

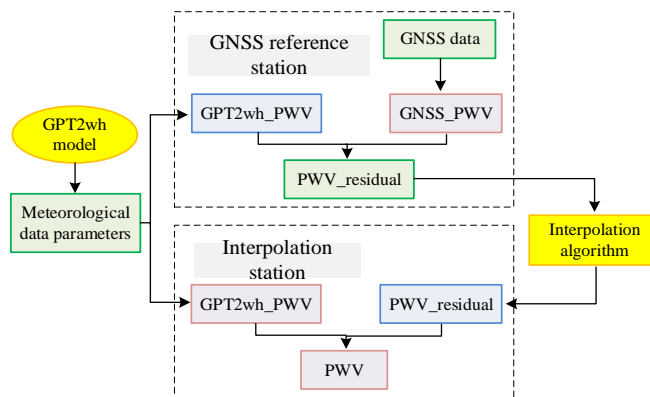


Fig. 2. Interpolation method process.

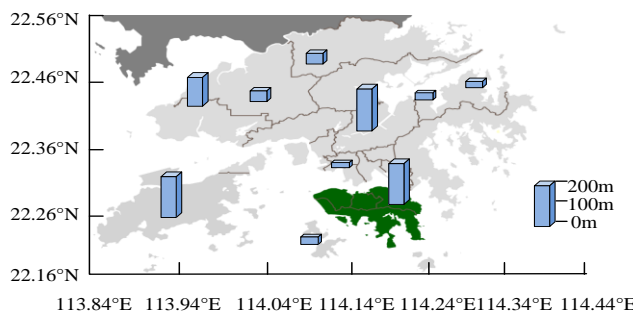


Fig. 3. Distribution and elevation of GNSS stations.

Based on the above content, research has estimated PWV in the atmosphere using the GPT2w model and Tm-Ts model. However, GNSS stations are often scattered and have large distances, which limits the analysis of AWWSD in some applications. Research conducts the spatial interpolation of GNSS-PWV to improve the accuracy of existing PWV interpolation methods.

2) *GA-based 3D water vapor AHP algorithm*: Through GNSS observation data and meteorological information, precise PWV of the station can be obtained, while PWV is only the water vapor content in a unit low area column that runs through atmosphere. It is the average water vapor content in the zenith direction of multiple rays on the oblique path near the station, and cannot accurately show the variation in water vapor space, thereby limiting its application in meteorology such as weather forecasting. A GNSS water vapor tomography (WVT) manner is put forward, which uses optimization methods for equation solving to avoid matrix inversion, but relies heavily on constraint equations and external observation data. The parameters of the water vapor chromatography method based on genetic algorithm are: the population size is 200, the crossover probability is set to 0.8, the Choice function is Roulette, the crossover function is Intermediate, and the mutation function is adaptive feasibility.

Chromatography techniques (CT) refer to the method of inverting the spatial distribution of parameters using integrated observations from different positions and directions within the study area. In GNSS 3D WVT, the oblique path water vapor content (SWV) generated by the GNSS satellite

signal passing through the tomography area is the observed measurement, and the WVD in each grid is the desired parameter. The WVT observation equation is established by calculating the intercept within each grid, as shown in equation (6).

$$SWV^q = \sum_{i=1}^n d_i^q \cdot x_i \quad (6)$$

In equation (6), q indicates the satellite ray number for WVT; n means the total number of tomographic grids; d_i^q means the intercept of the q th satellite ray passing through the i th grid, and x_i refers to the WVD value within the i tomographic grid. The equation diagram is shown in Fig. 4.

Due to the fixed position of GNSS stations in the chromatography area, satellite signals will vary with their altitude and azimuth angles. When constructing the observation equation for water vapor chromatography, these SWVs passing through the side of the chromatography study area are defined as invalid satellite rays, and only the effective satellite rays passing through the top of the chromatography area are selected. It collects all available WWVs for water vapor chromatography and constructs a water vapor chromatography equation set using equation (7).

$$y_{m \times 1} = A_{m \times n} \cdot x_{n \times 1} \quad (7)$$

In equation (7), y means the matrix containing all SWVs; m expresses the total number of SWVs that can be used for water vapor chromatography; A indicates the matrix

containing the intercept of satellite rays passing through the grid, where the number of never passing through the grid is 0. x means the matrix composed of all the parameters to be solved, and n means the total number of WVD in the grid to be solved. For some small-scale water vapor tomographic regions, the vertical layering of the tomographic grid is almost parallel, and the mutual exchange of water vapor densities in each layer of the grid may not affect the overall results, which leads to the multiplicity of observation equations. In addition,

the limited number of GNSS stations in the tomographic region often leads to uneven and insufficient distribution of GNSS satellite signal rays in the tomographic region. Furthermore, there is a case where the coefficient matrix A is rank deficient. Therefore, in addition to the observation equation, the study also constructed horizontal, vertical and top-level constraints. The geometric diagrams of the first two constraint equations are shown in Fig. 5.

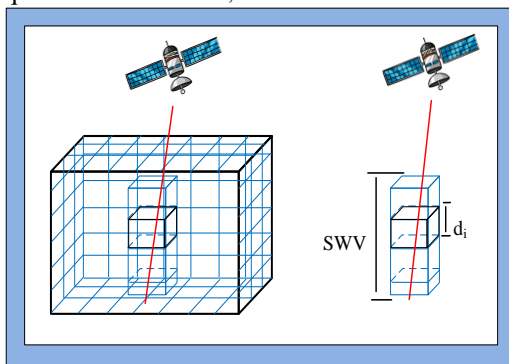


Fig. 4. Graphical representation of GNSS three-dimensional WVT observation equation.

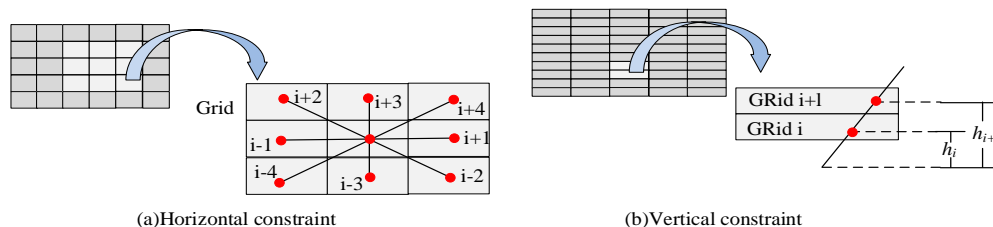


Fig. 5. Geometric diagram of horizontal and vertical constraint equations.

Equation (8) is established by using distance dependent Gaussian weighting function.

$$w_1^i x_1 + w_2^i x_2 + \dots + w_{i-1}^i x_{i-1} - x_i + w_{i+1}^i x_{i+1} + \dots + w_l^i x_l = 0 \quad (8)$$

In equation (8), l indicates the total amount of grids in the same horizontal layer; x_l denotes the atmospheric WVD of the l th grid; w_l^i expresses the horizontal weight coefficient of the l th grid in the horizontal layer relative to the i th grid, and d is the distance between the center points of the corresponding two grids. The vertical constraint equation is shown in equation (9).

$$x_i = x_{i+1} \cdot e^{(h_{i+1} - h_i)/h} \quad (9)$$

In equation (9), h is the height of the water vapor chromatography area. For top level constraints, research suggests that the top level region of the tomographic model has very little water vapor, so the WVD of all top level tomographic grids is directly constrained to be 0. Synthesizing three constraints and using the least square (LS) method can obtain the atmospheric WVD solution as shown in equation (10).

$$x = (A^T P A + B^T P B)^{-1} \cdot (A^T P y) \quad (10)$$

In equation (10), B denotes the coefficient matrix of the constraint equation. For the 3D WVT method based on GA, it needs to primarily construct the tomography equation, and then the idea of optimization equation is used to solve it. The calculation method is shown in equation (11).

$$\min f(x) = (y - Ax)^T P (y - Ax), x \in R^+ \quad (11)$$

In equation (11), the x value that minimizes $f(x)$ is the result of water vapor chromatography. After the tomographic equation is obtained, a fitness function is constructed, and then some groups representing the approximate value of the grid WVD are randomly generated. Then, according to the fitness value of the grid WVD, the grid density estimation value of the subsequent generations is selected as the parent. After selection, a new grid WVD approximation solution is formed by using the group of parents mentioned above to calculate new offspring through crossover and mutation. It calculates the fitness value of the approximate value of the WVD of each group of grids. When the fitness value of a group meets the requirements or reaches the number of searches, the GA algorithm stops searching, or continues to iterate circularly. The specific process of GA based water vapor chromatography method is shown in Fig. 6.

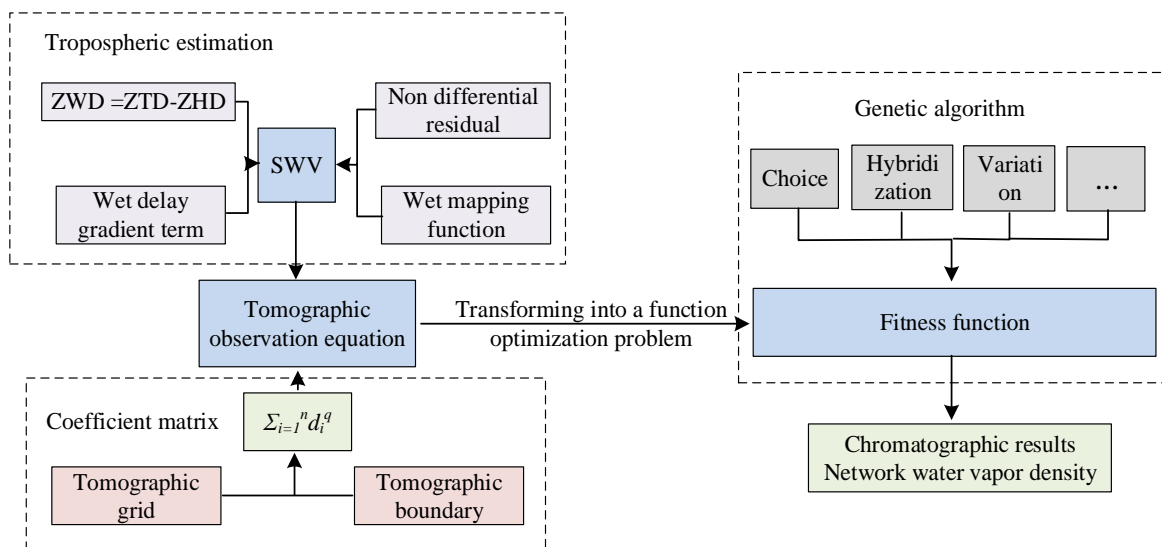


Fig. 6. Flow chart of water vapor chromatography method based on GA algorithm.

Based on the above operations, x that minimizes the value of the fitness function is obtained, that is, the best solution of equation (11), from which the distribution information of atmospheric WVD over the tomographic area is obtained.

IV. RESULTS AND DISCUSSION

The purpose of the performance evaluation is to measure the good or bad results of the research work in order to further identify and optimize the performance differences between different models. In order to achieve a more accurate solution of water vapor distribution, research will be carried out on the estimation of atmospheric precipitable water volume (PWV) based on GNSS technology, and the idea of removal interpolation restoration will be introduced to optimize the accuracy of the calculation. In terms of three-dimensional water vapor chromatography, a chromatography method based on genetic algorithm is proposed to solve the problems in the chromatography equation. To test the detection performance of the water vapor spatial distribution detection method developed by the research institute, a series of experiments were designed to test the model improvement effect and detection accuracy. When GPT2w model was applied to work out Tm, the study used a decreasing rate to adjust and optimize it vertically. Here, the optimized model was named GPT2wh. To test the improvement effect of the model, the accuracy of GPT2wh was verified by using Tm calculated from sounding station data. The experiment selected 459 sounding stations containing meteorological profile information for more than half a year in 2020, and used the two improved models before and after to obtain the daily Tm values of the corresponding stations from 0 to 12 o'clock. Bias and RMSE were used as statistical variables for comparative analysis, as shown in Fig. 7.

In Fig. 7, for the GPT2wh model, the accuracy of most sounding stations was better than 5K, with a proportion of 88.45%; For the GPT2wh model, stations with an accuracy of 5K accounted for 76.45%. Compared to the GPT2w, the maximum improvement of RSME by the GPT2wh model was

7.35K, from 11.35K to 4.00K. The mean RMSE of the GPT2wh model was 3.83K, which was 0.33K less than that of the GPT2w, and the accuracy was improved by about 8%. The different colors in Fig. 7(a) and 7(b) could distinguish between different types of stations, with positive and negative Bias stations, respectively. This was mainly due to the height difference between the sounding station and its four model grid points, while the GPT2wh model included a Tm decay rate that could be vertically corrected for Tm, with significant reductions in positive and negative bias values. The mean deviation of the GPT2wh model was 0.32K and the mean deviation of the GPT2w model was 0.94K. Based on the contents of Fig. 7, the improved model gave a more accurate Tm value.

To further test the performance of the two models, statistical analysis was conducted on the height difference between each sounding station and its corresponding model nodes, and the trend curve of the two indicators changing with the height difference was plotted in Fig. 8.

In Fig. 8, the RMSE and Bias values of the GPT2w increased with the increase of height distinguish, while the GPT2wh could get stable RMSE and Bias in different height difference ranges. The average bias of the GPT2w model varies widely at different altitudes, from 0.32k to 10.34k; the average bias of the GPT2wh model is relatively stable and small. As the altitude difference increases, the average RMSE of the GPT2w model increases significantly, while the GPT2wh model can achieve smaller and smaller RMSE. It can be seen that the GPT2wh model has limited improvement in Tm when the height difference is small, while the improvement effect on Tm is more significant when the height difference is large. This indicated that the improved model could well promote the accuracy of height difference method station estimation Tm. And as the height difference increased, the promotion of the GPT2wh became more significant.

To validate the effectiveness of the PWV spatial interpolation method proposed by the research institute, station cross validation and grid point PWV interpolation

using GNSS station PWV were used to compare the interpolation accuracy of the entire region. Firstly, different interpolation schemes were used for comparison: the first group of schemes did not take into account of the influence of elevation factors on PWV interpolation, and directly used IDW, Kriging, and TPS. The second group added the GPT2wh model on top of the first group, while the third group used the 3DRing and 3DTPS algorithms that considered elevation factors. Adding the GPT2wh model on top of the third group formed the fourth group of schemes. By comparing the PWV interpolation schemes mentioned above, the impact of elevation factors on the PWV interpolation results has been verified, and the effectiveness of using the removal interpolation restoration idea for PWV interpolation in different situations has been evaluated. The RMSE distribution of daily station cross validation for different schemes is shown in Fig. 9.

In Fig. 9, the first group of schemes showed the worst with a larger RMSE, while the third group considered the elevation factor and reduced the RMSE of the interpolation results. Based on the idea of removal interpolation recovery, after adding the GPT2wh model, the second group effectively reduced the RMSE value compared to the first group and the fourth group compared to the third group. Based on the content of Fig. 9, this idea could effectively improve the interpolation accuracy of PWV. In the results of Kriging and 3DKriging, the interpolation algorithm that adds elevation factors can improve the accuracy of PWV interpolation for most stations, but some stations do not show significant improvement, especially for stations located at 113.84° E

and 22.26° N. Because when the HKNP station is the desired station, its elevation is no longer within the elevation range of the other 11 reference stations, and the elevation is relatively large. Even if the influence of elevation factors is taken into account in 3DDrilling, 11 reference stations cannot provide sufficient and reliable elevation reference information for the HKNP station, resulting in a small improvement in PWV interpolation accuracy. The second highest station is HKST. Although its elevation is significantly different from most of the survey stations, its elevation is included in the elevation range of 11 reference stations. Therefore, using Kriging for PWV interpolation accuracy crossover, while using 3DKriging can better improve the PWV accuracy of the HKST survey station. Through analysis, it can be concluded that the accuracy of 3DDrilling interpolation of PWV depends largely on the selection of reference stations, often requiring the elevation range of reference stations to cover as much as possible the elevations of all stations to be measured. The same problem exists for 3DTPS. The PWV interpolation method, which is based on the idea of removal interpolation restoration, can effectively solve the above problems. For the HKNP station, the RMSE of Kriging GPT2wh interpolated PWV is 1.58m, which is 4.45m better than the RMSE of Kriging; the RMSE of TPS-GPT2wh interpolated PWV is 1.43m, which is 2.97m better than the RMSE of TPS. Compared to PWV interpolation using 3DDrilling and 3DTPS, Kriging-GPT2wh and TPS-GPT2wh not only improve the accuracy of HKNP for special height stations, but also perform better for all other stations.

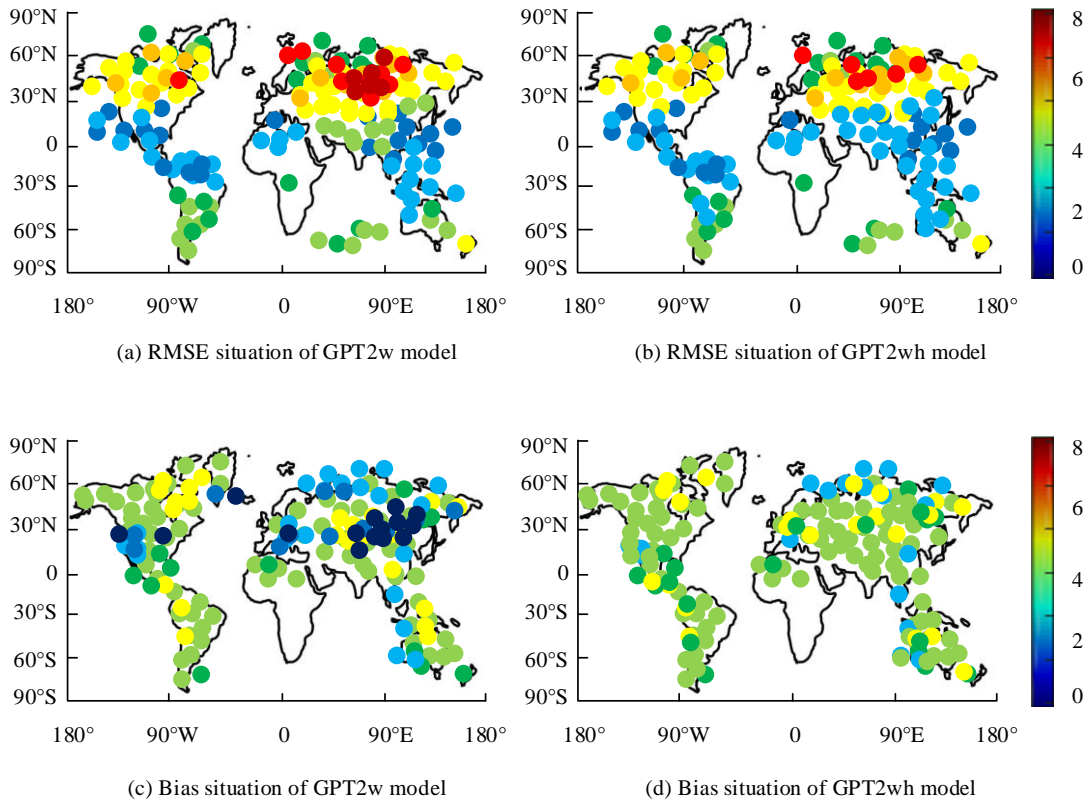


Fig. 7. Bias and RMSE scatter plots of the GPT2w model before and after improvement.

To further verify the accuracy of each interpolation method, the PWV of each grid point interpolated by ten methods was statistically compared with the PWV provided by ECMWF. The statistical information of grid points for different interpolation methods is listed in Table I.

In Table I, the grid point PWV verification results were similar to the station cross validation results. The accuracy of grid point verification results was better than that of station cross validation, regardless of MAE, RME, or CRE. This was because of the overall difference in PWV obtained from ECMWF grid data and GNSS data.

The LS method was the most commonly used solution method for 3D WVT, and a large amount of experiments demonstrated that this method could obtain high-precision atmospheric WVD. To assess the GA algorithm's accuracy based tomography method designed by the research institute, the results were compared with those of LS. In the experiment, linear regression and box chart were used to analyze the atmospheric water vapor density distribution of the two methods. The chart included experiments under all weather conditions, as shown in Fig. 10.

In Fig. 10 (a), the scatter distribution and regression line of atmospheric WVD showed a good linear relationship between the two chromatographic methods. The beginning point and

slope of the regression equation were 0.534 and 0.9542, respectively. Fig. 10(b) showed the distribution of the difference in atmospheric WVD between the two chromatography methods, with Q1 and Q3 being $-0.83\text{g}/\text{m}^3$ and $0.61\text{g}/\text{m}^3$, respectively, indicating that the proportion of the difference in atmospheric WVD calculated by the two chromatography methods within $1\text{g}/\text{m}^3$ exceeded 50%. The upper and lower limits of the box plot were $2.73\text{g}/\text{m}^3$ and $-2.88\text{g}/\text{m}^3$, respectively. The outlier only accounted for 3.21%. From the content of Fig. 10, the atmospheric density outcomes obtained by the GA based tomography method were consistent with that obtained by the LS method, which proved that this method could achieve high precision solution of AWVSD. The RMSE and MAE obtained by comparing the genetic algorithm based tomography results with sounding data and ECMWF data were $1.45/1.24\text{g}/\text{m}^3$ and $1.35/1.01\text{g}/\text{m}^3$, respectively, while the RMSE/MAE obtained by comparing the least squares tomography results with sounding data and ECMWF data were $1.46/1.24\text{g}/\text{m}^3$ and $1.37/1.15\text{g}/\text{m}^3$, respectively. Overall, it can be seen that both the genetic algorithm based tomography and the least squares tomography methods can provide good atmospheric water vapor results compared to the reference values in both sunny and rainy experiments, and the statistical results of the former are better than those of the latter.

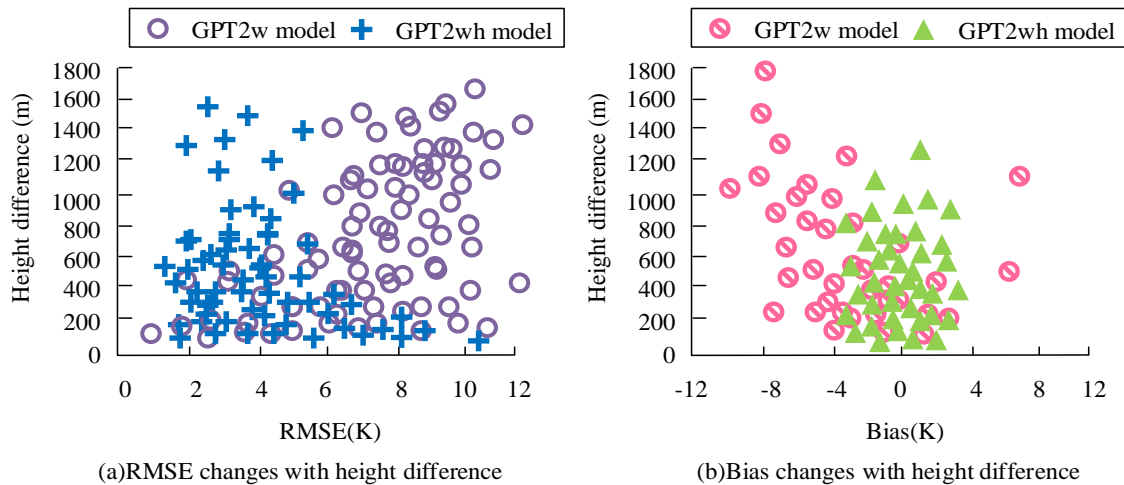


Fig. 8. Variation of RSME and bias for estimating t_m values by different models with station altitude difference.

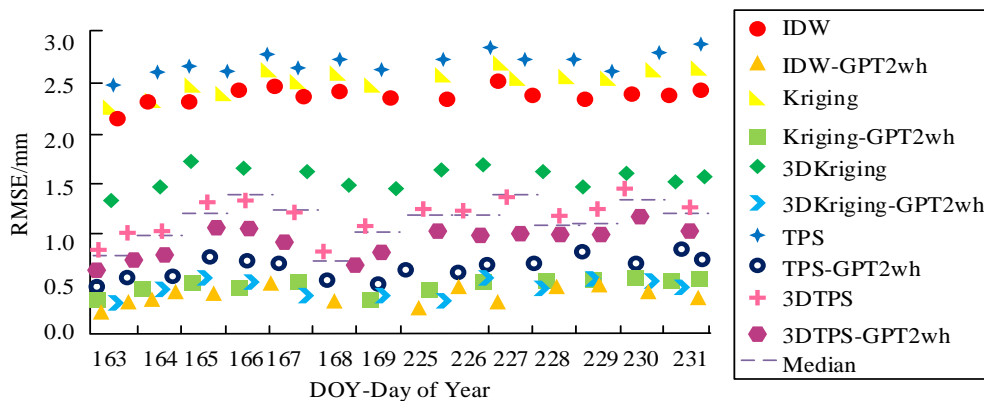


Fig. 9. RMSE distribution of various PWV interpolation methods in different year product days.

TABLE I. COMPARISON OF GRID POINT STATISTICAL INFORMATION OF DIFFERENT INTERPOLATION METHODS

Interpolation method	Sunny day(mm)			Rain(mm)		
	MAE	RMSE	CRE	MAE	RMSE	CRE
IDW	2.459	2.966	0.462	2.455	2.967	0.471
IDW-GPT2wh	1.640	2.045	0.193	1.643	2.045	0.196
Kriging	2.537	3.012	0.472	2.533	3.023	0.471
Kriging-GPT2wh	1.470	1.801	0.141	1.481	1.803	0.143
3DKriging	1.820	2.203	0.217	1.812	2.204	0.220
3DKriging-GPT2wh	1.587	1.896	0.160	1.592	1.892	0.162
TPS	2.873	3.324	0.692	2.882	3.312	0.693
TPS-GPT2wh	1.630	1.923	0.168	1.631	1.935	0.169
3DTPS	1.978	2.375	0.245	1.974	2.348	0.243
3DTPS-GPT2wh	1.699	2.004	0.182	1.698	2.003	0.181

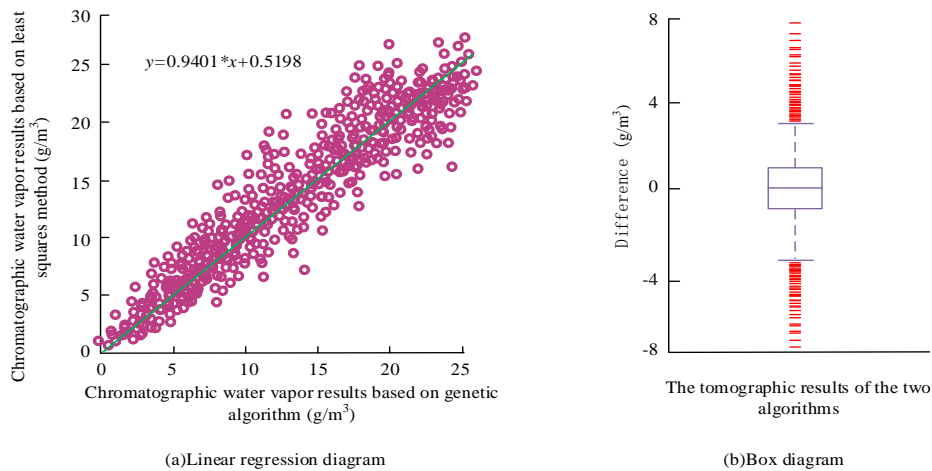


Fig. 10. Linear regression and box plot of chromatographic results obtained from two chromatographic methods.

To further test the performance of the tomography method (Method 1) designed by the research institute, three-dimensional WVT experiments were conducted under different weather conditions using GNSS observation data offered by the Hong Kong Satellite Positioning Reference Station Network SatRef, and three-dimensional water vapor information was obtained under different weather conditions. Compare Method 1 with the currently popular centralized chromatography method, and the specific results were shown in Table II. The comparison methods included: GNSS WVT with consideration for boundary signals and vertical constraints (Method 2), a robust adaptive WVT (Method 3), WVT with fusion of ECMWF grid data (Method 4), and water vapor distribution tomography based on Kalman filtering (Method 5).

In Table II, regardless of whether it is sunny or rainy, the RMSE values of water vapor chromatography density for Method 1 were lower than those for the other four methods. During the entire chromatographic experiment, the average RMSE and MAE values of Method 1 were 1.78g/m³ and 1.41g/m³, respectively. According to the comprehensive table, Method 1 could obtain higher precision atmospheric WVD values. The meteorological profile data provided by radiosonde stations can be used to calculate water vapor

density values, which are studied as reference values to evaluate the accuracy of water vapor tomography methods based on genetic algorithms. Due to the daily launch of radiosonde balloons at 0:00 and 12:00 UTC, the study chose to compare the water vapor chromatography results of the corresponding time periods under sunny and rainy conditions. During the rainy period from August 9, 2022 to August 15, 2022, the atmospheric water vapor density calculated from radiosonde data and the results of water vapor tomography based on genetic algorithms will vary with altitude. It is found that the atmospheric water vapor density decreases with increasing altitude; the atmospheric water vapor profile obtained by the genetic algorithm based tomography method is in agreement with the atmospheric water vapor profile obtained by radio sounding. In absolute terms, the agreement is better in the upper atmosphere. Considering the relative error, the corresponding values for tomographic grids with heights greater than 5km and less than 5km are 31% and 15%, respectively. This is because the water vapor density value in the upper atmosphere is relatively small, and small differences between sounding data and tomographic results can also lead to significant relative errors; the proportion of the atmospheric water vapor content within 5km of the Earth's surface exceeds 90%.

TABLE II. COMPARISON OF WATER VAPOR CHROMATOGRAPHY RESULTS OF VARIOUS METHODS

Project		Method 1	Method 2	Method 3	Method 4	Method 5
Sunny day	RMSE	1.84	2.01	1.98	2.22	2.31
	MAE	1.42	1.98	1.74	2.01	2.14
Rain	RMSE	1.71	2.03	1.96	2.19	2.27
	MAE	1.39	1.96	1.79	2.00	2.12
Average value	RMSE	1.78	2.02	1.97	2.21	2.29
	MAE	1.41	1.97	1.77	2.00	2.13

V. CONCLUSION

The amount of water content in clouds not only has a significant impact on the growth of cloud droplets and the formation and intensity of precipitation, but also serves as a predictive parameter for global climate data simulation and an important parameter for studying the impact of clouds on climate. To achieve a more accurate solution of AWVSD, research was conducted on estimating PWV based on GNSS technology, while introducing the idea of removal interpolation restoration to optimize the accuracy of the calculation. In three-dimensional water vapor chromatography, a chromatography method based on GA was proposed to address the problems in the chromatography equation. Through experimental analysis, the accuracy of most sounding stations in the GPT2wh model was better than 5K, with a proportion of 88.45%, which was more accurate than the Tm value obtained before improvement. The first group of schemes displayed the worst with a larger RMSE, while the third group considered the elevation factor and reduced the RMSE of the interpolation results. Based on the idea of removal interpolation recovery, after adding the GPT2wh model, the second group effectively reduced the RMSE value compared to the first group and the fourth group compared to the third group. This idea could effectively improve the interpolation accuracy of PWV. The average RMSE and MAE values for Method 1 were 1.78g/m³ and 1.41g/m³, respectively. The tomography method designed by the research institute could obtain higher precision atmospheric WVD values. The detection method for water vapor spatial distribution proposed by the research institute needs to be further improved and refined. Further research and work are needed after this proposal:

Step 1: Various empirical meteorological parameter models have been studied and constructed, but with the continuous enrichment and improvement of analytical data and spatial resolution of meteorological data, more dense meteorological grid parameter models need to be provided. In future research work, it is necessary to further consider how to reasonably sample global grid points, refine the positional relationship between the desired points and grid points, and improve the accuracy of spatial distribution research.

Step 2: A stable 3D water vapor chromatography algorithm and program have been developed, but the efficiency and automation level of the program still need to be improved. Subsequent research content needs to achieve automated batch processing of GNSS observation data, automated mapping,

and comparative verification of sounding data.

Step 3: The study area selected by the Research Institute is the Hong Kong region, which has a large number of GNSS stations and a small geographical area. However, the research has not explored how to achieve water vapor spatial distribution detection in areas with large areas and sparse GNSS stations. In the future, regional water vapor spatial inversion can be achieved by exploring reasonable grid partitioning strategies, multi-source observation data fusion and other means.

ACKNOWLEDGMENT

The research is supported by: R&D center of building energy saving engineering technology of Shandong Huayu University of Technology, No.: 2019-03.

REFERENCES

- [1] Guo R., Han D., Chen W., Dai L., Ji K., Xiong Q., Müller-Buschbaum P. Degradation mechanisms of perovskite solar cells under vacuum and one atmosphere of nitrogen. *Nature Energy*, 2021, 6(10): 977-986.
- [2] Schumacher D L, Keune J, Dirmeyer P, Miralles D G. Drought self-propagation in drylands due to land-atmosphere feedbacks. *Nature geoscience*, 2022, 15(4): 262-268.
- [3] Halldorsson V. National sport success and the emergent social atmosphere: The case of Iceland: *International Review for the Sociology of Sport*, 2021, 56(4):471-492.
- [4] Xing G, Deng C, Di J, Zhu H, Yu C. High-temperature behaviour of V2AlC powders under nitrogen atmosphere. *Ceramics International*, 2022, 48(10): 14424-14431.
- [5] Zheng N, Chai H, Chen L, Ma Y, Tian X. Snow depth retrieval by using robust estimation algorithm to perform multi-SNR and multi-system fusion in GNSS-IR. *Advances in Space Research*, 2023, 71(3): 1525-1542.
- [6] Shi Y, Jiang Z, Liu Z, Li L. A Lagrangian analysis of water vapor sources and pathways for precipitation in East China in different stages of the East Asian summer monsoon. *Journal of Climate*, 2020, 33(3): 977-992.
- [7] Huang L, Mo Z, Xie S, Liu L, Chen J, Kang C, Wang S. Spatiotemporal characteristics of GNSS-derived precipitable water vapor during heavy rainfall events in Guilin, China. *Satellite Navigation*, 2021, 2(1): 1-17.
- [8] Lee J K, Han H S, Chaikasetin S, Marron D P, Waymouth R M, Prinz F B, Zare R N. Condensing water vapor to droplets generates hydrogen peroxide. *Proceedings of the National Academy of Sciences*, 2020, 117(49): 30934-30941.
- [9] Tan J, Yu D, Li Q, Tan X, Zhou W. Spatial relationship between land-use/land-cover change and land surface temperature in the Dongting Lake area, China. *Scientific reports*, 2020, 10(1): 1-9.
- [10] Pan Y, Ren C, Liang Y, Zhang Z, Shi Y. Inversion of surface vegetation water content based on GNSS-IR and MODIS data fusion. *Satellite Navigation*, 2020, 1(1): 1-15.

- [11] Zheng N, Chen P, Li Z. Accuracy analysis of ground-based GNSS-R sea level monitoring based on multi GNSS and multi SNR. *Advances in Space Research*, 2021, 68(4): 1789-1801.
- [12] LÜ M Z, CHEN K J, CHAI H S, CENG J, ZHANG S, FANG L. Joint inversion of InSAR and high-rate GNSS displacement waveforms for the rupture process of the 2022 Qinghai Menyuan M6. 9 earthquake. *Chinese Journal of Geophysics*, 2022, 65(12): 4725-4738.
- [13] Lewen Z, Jiaqian R, Yang D. Platform for GNSS real-time space environment parameter inversion and its accuracy evaluation. *Nanjing Xinxing Gongcheng Daxue Xuebao*, 2021, 13(2): 204-210.
- [14] Jalali Z, Noorzai E, Heidari S. Design and optimization of form and facade of an office building using the genetic algorithm. *Science and Technology for the Built Environment*, 2020, 26(2): 128-140.
- [15] Garud K S, Jayaraj S, Lee M Y. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. *International Journal of Energy Research*, 2021, 45(1): 6-35.
- [16] Shariati M, Mafipour M S, Mehrabi P, Ahmadi M, Wakil K, Trung N T, Toghrol A. Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm). *Smart Structures and Systems, An International Journal*, 2020, 25(2): 183-195.
- [17] Abualigah L, Alkhrebsheh M. Amended hybrid multi-verse optimizer with genetic algorithm for solving task scheduling problem in cloud computing. *The Journal of Supercomputing*, 2022, 78(1): 740-765.
- [18] Yang Y, Song X. Research on face intelligent perception technology integrating deep learning under different illumination intensities. *Journal of Computational and Cognitive Engineering*, 2022, 1(1): 32-36.
- [19] Guo Y, Mustafaoglu Z, Koundal D. Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms. *Journal of Computational and Cognitive Engineering*, 2023, 2(1): 5-9.
- [20] Hidayat I, Ali M Z, Arshad A. Machine Learning-Based Intrusion Detection System: An Experimental Comparison. *Journal of Computational and Cognitive Engineering*, 2022, 2(2):88-97.

Detection of Tuberculosis Based on Hybridized Pre-Processing Deep Learning Method

Mohamed Ahmed Elashmawy¹, Irraiyan Elamvazuthi², Lila Iznita Izhar³, Sivajothi Paramasivam⁴, Steven Su⁵

Smart Assistive and Rehabilitative Technology (SMART) Research Group,

Department of Electrical and Electronic Engineering,

Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 32610, Malaysia^{1, 2, 3}

School of Engineering, UOWMKDU University College, Shah Alam, 40150, Malaysia⁴

School of Biomedical Engineering, University of Technology Sydney, Ultimo, 2007, Australia⁵

Abstract—The disease, tuberculosis (TB) is a serious health concern as it primarily affects the lungs and can lead to fatalities. However, early detection and treatment can cure the disease. One potential method for detecting TB is using Computer Aided Diagnosis (CAD) systems, which can analyze Chest X-Ray Images (CXR) for signs of TB. This paper proposes a new approach for improving the performance of CAD systems by using a hybrid pre-processing method for Convolutional Neural Network (CNN) models. The goal of the research is to enhance the accuracy and Area Under Curve (AUC) of detection for TB in CXR images by combining two different pre-processing methods and multi-classifying different manifestations of the disease. The hypothesis is that this approach will result in more accurate detection of TB in CXR images. To achieve this, this research used augmentation and segmentation techniques to pre-process the CXR images before feeding them into a pre-trained CNN model for classification. The VGG16 model managed to achieve an AUC of 0.935, an accuracy of 90% and a 0.8975 F1-score with the proposed pre-processing method.

Keywords—Tuberculosis; CNN; pre-processing; CXR images; augmentation; segmentation

I. INTRODUCTION

Tuberculosis, also known as TB, is an infectious disease brought on by the bacterium *Mycobacterium tuberculosis*. Although it mostly affects the lungs, it can also have an impact on the kidneys, the spine, and the brain.

When an infected individual coughs, sneezes, or talks and another person inhales the bacteria, TB is transmitted through the air. It is crucial to remember that TB cannot be transmitted via innocuous touch such as handshakes or sharing of utensils. Chronic cough, chest pain, blood in the cough, exhaustion, fever, night sweats, and weight loss are some of the signs of TB. Yet, some TB patients may not even exhibit any symptoms [1]. Although TB is a treatable and curable illness, a lengthy antibiotic treatment regimen is necessary. TB can be fatal if neglected.

According to the World Health Organization (WHO), tuberculosis (TB) is one of the top 10 causes of death worldwide. In 2020, there were an estimated 10 million new cases of TB globally with an estimated 1.5 million deaths from TB in 2020. The global TB treatment success rate was 85% in 2019 and disproportionately affects vulnerable populations,

such as people living in poverty, people who use drugs, and prisoners [2].

Chest X-rays and CT (Computed Tomography) scans are two different medical imaging techniques used to diagnose and monitor various conditions related to the chest, such as lung diseases. However, considering the disproportionality of TB affecting impoverished areas, chest X-rays have the advantage regarding its cost, availability, convenience and quick results [3].

Sputum microscopy, Chest X-rays (CXR), and culture in solid and liquid media can all be used to diagnose and find TB. CXR, one of the most popular and cost-effective imaging tests worldwide, can be utilized for TB early detection [3]. Although CXRs are helpful for making an early diagnosis, radiologists may encounter difficulties, such as the inability to tell TB from other symptoms in some situations [4], [5].

Computer aided diagnosis (CADx) and computer aided detection (CADE) systems also known as CAD systems, have been shown to improve the accuracy of medical diagnoses, reduce false positives and false negatives, and potentially reduce the time needed for interpretation. However, like any diagnostic tool, CADE/x systems are not infallible and must be used in conjunction with clinical judgment and expertise. Overall, CADE/x systems are a valuable tool in modern medicine, helping medical professionals to make more accurate diagnoses and improve patient outcomes [6]. A CAD system traditionally consisted of four main stages; pre-processing, segmentation, feature extraction and classification [7].

To optimize the system and produce higher accuracy, a wide variety of techniques and algorithms can be applied at each step, in various combinations [8]. Classifiers make the final determination regarding the patient's health state among the four stages of a CAD system. Information from earlier stages is compressed and filtered to acquire the information that is most pertinent to the patient's health and is then fed to classifiers [9]. Machine learning emerged in the field of computer science that enabled computers to classify data without being explicitly programmed. As the field of machine learning (ML) research developed, many classification algorithms, including Decision Trees, Support Vector Machines (SVM), Genetic Algorithms (GAs), and Fuzzy Algorithms (FA), flourished [10-12]. However, Convolutional

Neural Networks (CNNs), perhaps more significantly, have recently demonstrated their reliability [13].

CNN is a deep learning technique which makes it ideal for big data. The arrangement of the visual cortex in the brain served as the model for CNN construction. It is made up of multiple layers of linked neurons that work in a hierarchical fashion to interpret information ranging from basic features like edges and corners to more intricate forms and patterns. [14].

The objective of this research is to investigate the effects of combining two main methods of pre-processing methods such as shear, zoom and flipping for augmentation and the robustness of U-Net mask for segmentation on the training set of a pre-trained CNN classifier in hopes to a high accuracy and AUC with the proposed method of pre-processing on binary and multi-classification. Hence, being able to classify different types of manifestations of TB in CXR images.

The literature review is discussed in Section II. Section III describes the chest X-ray dataset that is used and the proposed methodology for pre-processing, the performance measurements are also explained in this section. Section IV discusses the experiments conducted and results obtained by applying the proposed methodology on the CXR dataset and the comparison with other papers. Finally, the proposed research work is concluded.

II. RELATED WORK

To identify any lung-related disease on CXRs, Computer Assisted Diagnosis (CAD) has been extensively used. Methods involving machine learning alternatives to CNN have been researched as CAD systems. Some methods of TB detection make use of the segmentation of the lungs. Here, scientists attempt to separate the heart or lung structures before assessing them for any anomalies. Other research implements augmentation methods onto the CXR images, changing the parameters of the images within a designated range before feeding it into the machine learning model to achieve more reliable results.

Related works show research on other models used aside from CNN with pre-processing for detection of TB in CXR images. Antony B. et al. [15] attempted eliminating background noise by applying two segmentation methods, the CANNY algorithm and a median filter on 662 images. The results reported only an 80% accuracy as the highest accuracy while using a K-NN classifier in his paper, the highest accuracy among the other two classification methods (SMO and SLR).

To detect Tuberculosis in patients, Muhathir [16] used the K-NN classification method and HOG feature extraction technique. The results indicate that 70.90% of positive cases were correctly identified, with 234 out of 330 samples, while 72.72% of negative cases were correctly identified, with 240 out of 330 samples. The study shows that using the K-NN and HOG feature approach, the X-ray Set TB can be classified with an accuracy of 71.81% when using cross-validation.

Three alternative deep-learning models—AlexNet, ResNet-18, and DenseNet121—were tested in Jared et al. [17] study to

see which was most effective in identifying tuberculosis (TB) in CXR. For their training set, the researchers used 180,000 images, but they made no mention of data pre-processing. According to the findings, DenseNet121 had the highest accuracy (91%), with an Area Under Curve (AUC) ranging from 0.94 to 0.96. Although Jared et al. [17] did not rely on augmented photos to increase the size of the dataset, the high accuracy and AUC were probably caused by the numerous training images employed.

For the purpose of TB identification in CXR pictures, Syeda et al. [18] developed an ensemble model using the pretrained deep-learning models VGG-16, VGG-19, ResNet50, and GoogleNet. The ensemble's accuracy was 86.7% with an AUC of 0.92 after training on 600 images and 200 more; however, if any pre-processing techniques were used, the accuracy might have been enhanced with a bigger data set and variance to prevent overfitting problems.

Gordienko et al. [19] reported an increase in accuracy and loss after segmentation of 247 TB CXR images with a U-Net CNN and Bone Shadow Exclusion and using them for training a self-made seven-layer CNN. No reports of augmentation have been applied to the image, and with a low number of images overfitting could occur. It has only been reported that the test accuracy has increased, and the test loss has decreased.

Erdal [20] proposes and compares in his study, three methods of segmentation: bounding box, lung mask with black background and lung mask with white background. AlexNet, VGG16 and VGG19 deep-learning architecture were utilized for feature extraction and a Random Forest algorithm for classification. Accuracy reached 88.3% and AUC reached 0.93 as reported. Erdal [20] also reported that the lack of contrast enhancements and augmentation have acted as a limitation in his research.

Only using the Montgomery County (MC) TB CXR images dataset, with 138 images total of both abnormal and normal CXR images, Mustapaha & Serestina [21], have managed to multiply the total images up 5000 images through augmentation. Through their proposed CNN model a 87.1% accuracy was achieved.

Oposing previous studies that have utilized segmentation as a pre-processing technique for lung segmentation before feeding the images to the training model, such as the studies done by Erdal [20] and Gordienko et al. [19], Ahsan et al. [22] displayed that it is possible to achieve a comparable accuracy using the VGG-16 model without pre-processing segmentation. An accuracy of 80% was reached by the VGG-16 model and an accuracy of 81.25% when partial augmentation was applied.

Marcio et al. [23] combined the Montgomery, Shenzhen and PadChest CXR datasets with a total of 290 images, generating the training and test datasets using a HDF5 dataset generator, then applied augmentation. Marcio et al. [23] tested out three different pretrained CNN models being AlexNet, GoogleNet and ResNet50, achieving results between 0.78 and 0.84 AUC.

Similarly, Eman et al. [24] has used Montgomery and the Shenzhen Datasets and has used augmentation to multiply the dataset images up to 2040 images. However, Eman et al. [24]

have explored a more specialized and advanced array of CNN models; ConvNet, Exception, ResNet50, VGG16 and VGG19. All CNN models have achieved an accuracy above 87% and a maximum of 90% and an AUC of 0.91.

To summarize the presented related work, it can be concluded that there is emphasis on augmentation for improving classification models, while sometimes in other related works there is neglect of preprocessing. However, the benefits of segmentation in combination with augmentation remain unexplored as it is deemed more time-consuming and difficult, particularly when dealing with big or complicated data sets [22]. Addressing this underutilization of segmentation alongside augmentation is key to unlocking their full potential for classification performance.

The main contribution of this research is to develop a hybrid pre-processing method to enhance data quality by reducing noise and outliers, while normalizing data for more effective CNN learning. This study demonstrates that combining segmentation and augmentation as pre-processing enhances model accuracy.

III. METHODOLOGY

A. Datasets

The data sets that will be used during the experimental stage are all obtained from online open-source image databases:

1) *Shenzhen dataset*: The dataset was collected in collaboration with Shenzhen No. 3 People's Hospital, Guangdong Medical College, Shenzhen, China. It contains 662 cases of chest X-rays, including 326 normal cases and 336 tuberculosis cases [25].

2) *Montgomery (MC) dataset*: The dataset was collected from the Department of Health and Human Services in partnership with Montgomery County, Maryland in the United States. The group consisted of 138 frontal chest radiographs from the Montgomery County Tuberculosis Screening

Program, of which 80 were normal and 58 were tuberculosis [25].

The datasets were divided into training/validation dataset and test dataset with a respective ratio of 80:20.

B. Flowchart

The flowchart in Fig. 1 outlines the process of training a VGG16 CNN model for image classification. The training dataset is pre-processed with ZCA, normalization, U-net segmentation, and augmentation techniques in that specific order before being used to train the model. The validation dataset is used to monitor model performance during training, and the test dataset is used for prediction and calculating model performance metrics.

C. Data Pre-Processing

Image resizing to 227x227, Mean Normalization and Standardization were applied to the set before segmentation and augmentation. Regarding segmentation, U-net CNN segmentation is being utilized for the pre-processing segmentation.

1) *Normalization and ZCA*: Normalization is used to scale the pixel values of an image to a range between 0 and 1. This is done to ensure that the pixel values are in a consistent range and to prevent the dominance of certain pixel values. Normalization is often done by dividing each pixel value by the maximum pixel value in the image [26]. ZCA is used to remove the correlation between the different color channels in an image. This is important because the color channels may be correlated, which can lead to redundant information and increased computational complexity. ZCA whitening transforms the image data so that each pixel value is uncorrelated with every other pixel value. This is done by performing eigenvalue decomposition of the covariance matrix of the pixel values and then transforming the data using the eigenvectors [27]. Both normalization and ZCA are useful for improving the performance of machine learning models on image datasets.

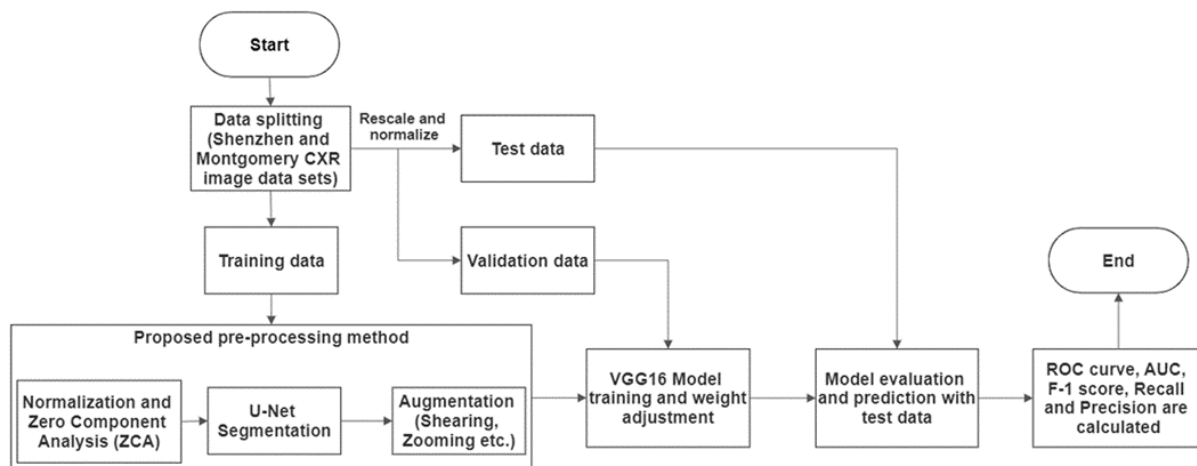


Fig. 1. Flowchart of simulation including the proposed method.

2) *U-net segmentation*: U-net CNN is a CNN designed for segmentation of biomedical images. What makes U-net stand out from other general convolutional neural networks is that general CNNs need to focus on image classification in biomedical cases, hence they require the user to identify the presence of a disease and localize the area of said disease. U-net eliminates this issue by applying classification to each pixel and can distinguish and localize borders by itself [28]. This method of segmentation could help eliminate the need for manual segmentation.

An expanding path and a contracting path make up the U-Net architecture. The contracting path is composed of convolutional layers followed by max-pooling layers, much like a conventional convolutional neural network. This path's goal is to minimize the spatial dimensionality of the input image while capturing its context.

The segmented image is localized using the expanding path, and the low-resolution feature maps created by the contracting path are upsampled. This path concatenates the corresponding feature maps from the contracting path with transposed convolutions to increase the spatial dimensionality of the features. This process is repeated several times to recover the original resolution of the input image.

The U-Net architecture also incorporates skip connections, which connect feature maps from the contracting path to feature maps from the expanding path. These connections aid in protecting the input image's small features, which can be lost during down sampling. That is where it gets its U shape from as seen in Fig. 2.

Many image segmentation tasks, including medical picture segmentation, cell segmentation, and object detection, have proven to be successful when using U-Net. It works particularly effectively in cases when the input images are tiny and the borders of the items that need to be divided are clearly defined [28].

3) *Augmentation*: Several transformations of augmentation were applied to the segmented images to increase the dataset's variability. The following augmentations were performed:

- Rescale transform every pixel value from range [0,255]
- Random rotations were applied up to 0.2 radians.

- Random shifts were applied to the width and height dimensions up to 0.1.
- Random shearing was applied for a maximum shear of 0.2.
- Zoom range up to 0.2.
- Horizontal and vertical flips.

D. VGG16 CNN Classifier

The VGG-16 architecture consists of 16 layers, including 13 convolutional layers, and three fully connected layers. It uses small 3x3 convolutional filters with a stride of 1 pixel, and max pooling layers with a 2x2 filter and stride of 2 pixels, which helps reduce the spatial dimensionality of the features.

The output of the last convolutional layer is flattened and fed into the fully connected layers, which perform classification on the input image. The final layer uses SoftMax activation to produce a probability distribution over the image classes.

Object detection, picture segmentation, and style transfer are just a few of the computer vision applications for which the VGG-16 architecture has been extensively employed as a pre-trained model [29]. Its popularity is a result of its efficiency and simplicity, which make it a reliable benchmark model for comparison with other designs.

1) *Fine-tuning*: The VGG-16 CNN model's weights are pretrained on a massive image dataset known as ImageNet. ImageNet is mostly comprised of natural and colorful pictures of animals and food, which deviates from what monochromatic Chest X-rays are. Fine-tuning the pretrained weights of the VGG-16 can be accomplished by unfreezing the weights of the 5th block of layers of the model. The unfreezing will increase the computation time needed for training the model, but it might result in more optimized weights [30].

2) *ADAM optimizer*: In order to calculate the difference between a neural network's predicted and actual output during training, the Adaptive Moment Estimation (ADAM) optimizer utilizes a loss function [31]. The optimizer then modifies the neural network's weights in accordance with the gradient of the loss function relative to the weights. Table I displays the hyperparameters used during experimentation.

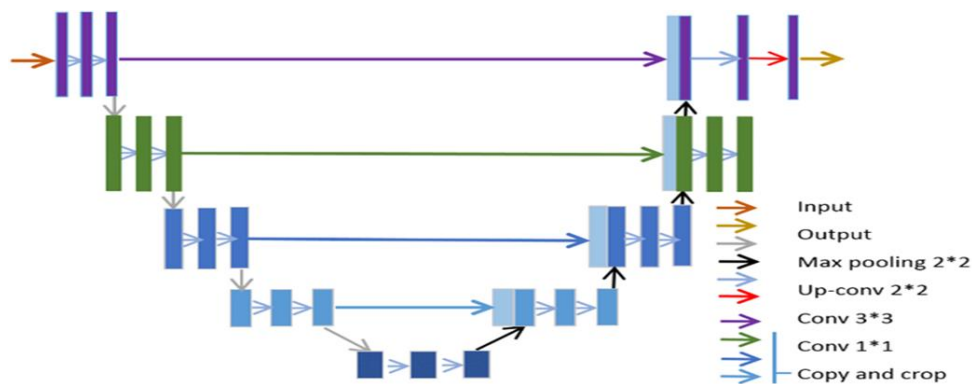


Fig. 2. The U-Net architecture displayed with the arrows denoting different operations [28].

TABLE I. HYPERPARAMETERS USED DURING EXPERIMENTATION

Hyperparameter	Values considered
Batch size	32,64
Number of epochs	30,40,50
Learning rate	0.001, 0.0001, 0.00001
Embedded dropout rate	0.3,0.5,0.7
Embedded activation function	Sigmoid
Optimizer	ADAM
Number of unfrozen layers	4 layers (5 th block)

E. Performance Evaluation Metrics

1) *Confusion matrix*: A confusion matrix is a table used to evaluate the performance of a classification algorithm by comparing the predicted and actual values of a dataset as seen in Table II. The matrix is organized into rows and columns, with each row representing the instances in a predicted class, and each column representing the instances in an actual class. The diagonal of the matrix represents the instances that were correctly classified, while the off-diagonal elements represent the instances that were misclassified.

The confusion matrix provides a useful way to visualize the performance of a classifier, and it can be used to calculate various metrics such as accuracy, precision, recall, and F1 score.

TABLE II. CONFUSION MATRIX

Actual values	Predicted values	
	Yes (Predicted class)	No (Predicted class)
Yes (Actual class)	True Positive (TP)	False Negative (FN)
No (Actual class)	False Positive (FP)	True Negative (TN)

- True Positive (TP): Number of patients correctly classifies as having TB.
- True Negative (TN): Number of patients correctly classified as not having TB.
- False Positive (FP): Number of patients incorrectly classified as having TB.
- False Negative (FN): Number of patients incorrectly classified as not having TB.

2) *Accuracy*: Accuracy of the model can be calculated from the confusion matrix by the following mathematical equation:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

3) *Precision*: Precision measures how many of the model's positive predictions were right. A high precision indicates that the model made few incorrect positive predictions and is a good indicator of how well the model detects positive cases. The precision mathematical equation is as follows:

$$P = \frac{TP}{TP + FP} \quad (2)$$

4) *Recall*: The model's recall evaluates how successfully it detects positive cases out of all actual positive cases. A high recall indicates that the model is good at recognising positive examples, whereas a low recall indicates that the model is missing many positive cases. The recall mathematical equation is as follows:

$$R = \frac{TP}{TP + FN} \quad (3)$$

5) *F1 score*: The F1-score is a measure of the balance between precision and recall. The F1-score is a useful metric for evaluating the overall performance of a binary classification model, especially when the classes are imbalanced. The F1-score equation is as follows:

$$F1 = \frac{2(Recall * Precision)}{(Recall + Precision)} \quad (4)$$

6) *ROC curve and AUC*: An ROC (Receiver Operating Characteristic) curve is a plot used to visualize the performance of a binary classification model. It is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at different classification thresholds, as seen in Fig. 3.

AUC (Area under the ROC Curve) is a valuable metric for evaluating the performance of binary classification models and is commonly used in a variety of machine learning applications. AUC represents the area under this curve, which ranges from 0 to 1. A perfect classifier would have an AUC of 1, while a random classifier would have an AUC of 0.5.

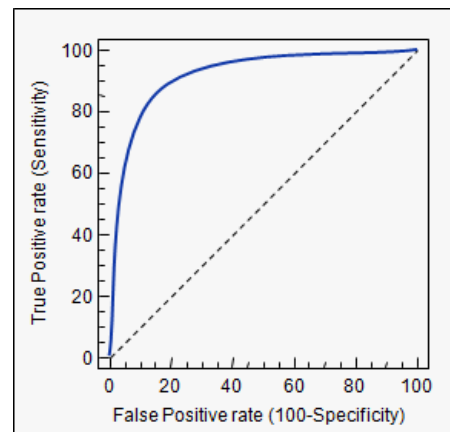


Fig. 3. ROC curve - A plot of true positive rate vs false positive rate.

IV. RESULTS AND DISCUSSION

A. Experiments

A total of three experiments have been conducted in this study to observe the effects of the hybridization of the pre-processing methods, such as no segmentation and augmentation (experiment 1), with segmentation only (experiment 2) and with augmentation and segmentation (experiment 3), where the results are discussed in terms of the confusion matrix and the other performance measures

discussed in Section III. The test data set consists of 160 CXR images with 80 TB positive cases and 80 normal cases.

B. Proposed Pre-Processing Method Performance

The hybridized pre-processing method used in experiment 3 has achieved a higher performance in comparison to experiment 1 and 2, as seen in the metrics achieved in the confusion matrix in table 3 and the performance measures in table 4, with an accuracy of 90%, a recall of 87.5%, a precision of 92.11% and a F1-score of 0.8975. Regarding the AUC shown in Fig. 4, out of the 3 experiments the proposed method has produced the best ROC curve and an AUC of 0.935, followed by 0.89 from experiment 2 and 0.87 from experiment 3, as seen from the confusion matrix table in Table III and the performance metrics table in Table IV.

C. Result Discussion

In this section, the discussion will progress from a baseline with no preprocessing to segmentation alone, and finally to the combined use of both techniques, the study systematically highlights their individual and joint impacts on model accuracy. This approach effectively highlights the research's

focus on demonstrating the effectiveness of segmentation and the combination of both segmentation and augmentation.

The first experiment has an accuracy of 82.5% and an AUC of 0.87 and does not use segmentation or augmentation. For the performance of the model, this approach is regarded as the standard. The input data is supplied straight to the CNN model without segmentation or augmentation, which may cause the data to be overfitted or underfitted.

The accuracy and AUC of the second experiment, which only uses segmentation, are 86.25% and 0.89, respectively. This method significantly outperforms the baseline method in terms of accuracy, proving that segmentation can increase the quality of the input data used to train the CNN model.

The third experiment involves both segmentation and augmentation and achieved 90% accuracy and an AUC of 0.935. This method shows the highest accuracy and AUC compared to the other two experiments. The combination of both segmentation and augmentation techniques could have helped the model to isolate important features and improve the robustness of the model.

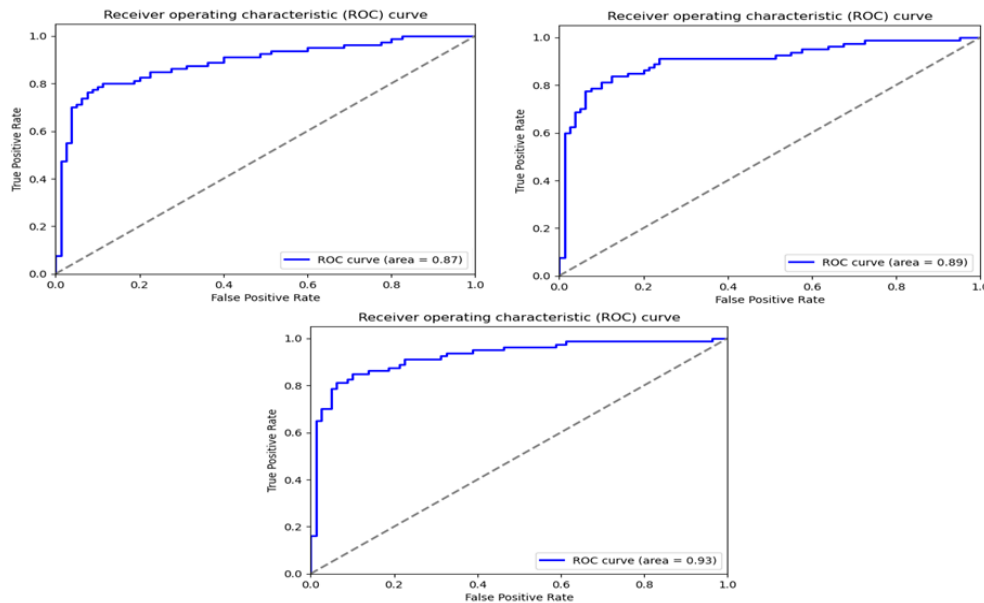


Fig. 4. ROC curves of experiment 1 (top left), experiment 2 (top right) and experiment 3 (bottom middle).

TABLE III. THE CONFUSION MATRIX OF EXPERIMENT 1-3 REPRESENTED IN NUMBER OF IMAGES

Experiment No.	Confusion matrix			
	True positive	True negative	False positive	False negative
1	56	76	4	24
2	64	74	6	16
3	70	74	6	10

TABLE IV. PERFORMANCE MEASURES CALCULATED FROM THE CONFUSION MATRIX

Experiment No.	Performance measures			
	Accuracy	Recall	Precision	F1-score
1	82.5%	70%	93.33%	0.8000
2	86.25%	80%	91.43%	0.8533
3	90%	87.5%	92.11%	0.8975

TABLE V. OVERALL CLASSIFICATION ACCURACY AND AUC COMPARISON WITH PREVIOUS RESEARCH WORK

Reference	CNN model	Dataset	Pre-processing	Overall Accuracy Rate and AUC
Proposed pre-processing method	VGG-16	Shenzhen and Montgomery	Augmentation and U-Net segmentation	90% and 0.935
Syeda et al. [18]	Ensemble (VGG-16, VGG-19, ResNet50 and GoogleNet)	Shenzhen and Montgomery	None	86.7% and 0.92
Ahsan M et al. [22]	VGG-16	Shenzhen and Montgomery	Augmentation	81.25% and 0.89
Erdal [20]	Ensemble (VGG-16, VGG-19, AlexNet, Random Forest)	Shenzhen	Manual segmentation	88.3% and 0.92

D. Comparative Analysis

The overall accuracy rate and AUC comparison of the various pre-processing methods from previous work is represented in Table V. For a valid comparison, the author has compared the proposed pre-processing method with other applications of pre-processing from the related works in Section II. It's crucial to maintain the experimental parameters as consistently as feasible.

The proposed pre-processing method, which included VGG-16, augmentation, and U-Net segmentation, achieved the maximum accuracy of 90% and AUC of 0.935, illustrating the efficiency of these techniques in improving the model's ability to generalize and detect essential traits in CXR images. Syeda et al.'s [18] technique showed lower accuracy and AUC even when employing multiple architectures, demonstrating that augmentation and segmentation are still beneficial and necessary in improving model performance.

As demonstrated by Ahsan M. et al.'s [22] method, augmentation alone was unable to improve model performance with a small data set size. While manual segmentation is useful as shown by Erdal [20], it is time-consuming and error-prone, and U-Net's ability to segment major features in images makes it more effective and robust for CAde systems.

V. CONCLUSION

In conclusion, the results of the study have shown that applying both segmentation and augmentation techniques can lead to better performance measures in terms of accuracy, AUC, recall, precision, and F1-score when classifying TB related CXR images compared to using only one or none of these techniques.

Augmentation has become a popular technique in recent years, and for good reason. It allows for the creation of a large and diverse training set without the need for additional data collection efforts, which can be time-consuming and costly. However, augmentation alone may not always be sufficient, particularly when dealing with complex images with intricate features or objects.

Furthermore, while manual segmentation can be effective in capturing important details in the images, it is not robust or flexible enough to be widely used in computer-aided detection systems. The rise of automatic segmentation methods such as U-Net has made segmentation more viable as a pre-processing technique for image analysis tasks. Overall, the findings suggest that researchers consider using both segmentation and augmentation techniques as pre-processing methods when developing CAde systems.

ACKNOWLEDGMENT

The authors would like to thank Ministry of Higher Education (MOHE), Malaysia and Universiti Teknologi PETRONAS for supporting this work through a Research Grant, FRGS, (Ref: FRGS/1/2022/TK07/UTP/02/4) and Cost Centre, (015MA0-159).

REFERENCES

- [1] Tuberculosis (TB) | American Lung Association. (n.d.). Retrieved December 8, 2019, from <https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/tuberculosis/>
- [2] W. H. Organization et al., "Global tuberculosis report 2020", World Health Organization, 2020.
- [3] Jaeger, S., Karargyris, A., Candemir, S., Siegelman, J., Folio, L., Antani, S., & Thoma, G. (2013). Automatic screening for tuberculosis in chest radiographs: a survey. *Quantitative Imaging in Medicine and Surgery*, 3(2), 89–99. <https://doi.org/10.3978/j.issn.2223-4292.2013.04.03>.
- [4] Ryu, Y. J. (2015, April 1). Diagnosis of pulmonary tuberculosis: Recent advances and diagnostic algorithms. *Tuberculosis and Respiratory Diseases*, Vol. 78, pp. 64–71. <https://doi.org/10.4046/trd.2015.78.2.64>.
- [5] Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. (2018) Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* 15(11): e1002686.
- [6] Li, Q., & Nishikawa, R. M. (Eds.). (2015). *Computer-aided detection and diagnosis in medical imaging*. Taylor & Francis.
- [7] Afsaneh Jalalian, Syamsiah B.T. Mashohor et al., Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review, *Clinical Imaging*, 37, 3, 5 2013.
- [8] Dheeba, J., Singh, N. A., & Selvi, S. T. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of biomedical informatics*, 49, 45-52.
- [9] Cascianelli, S., Scialpi, M., Amici, S., Forini, N., Minestrini, M., Luca Fravolini, M., ... & Palumbo, B. (2017). Role of artificial intelligence techniques (automatic classifiers) in molecular imaging modalities in neurodegenerative diseases. *Current Alzheimer Research*, 14(2), 198-207.
- [10] Sharon, H., Elamvazuthi, I., Lu, C.K., Parasuraman, S., Natarajan, E., 2019. Development of Rheumatoid Arthritis Classification from Electronic Image Sensor using Ensemble Method, *Sensors* 20, 167.
- [11] Timothy Ganesan, Pandian Vasant, Irraivan Elamvazuthi, *Advances in Metaheuristics, Applications in Engineering Systems*, CRC Press, 9 Dec, 2016.
- [12] Ku Abd. Rahim, K.N.; Elamvazuthi, I.; Izhar, L.I.; Capi, G. Classification of Human Daily Activities Using Ensemble Methods Based on Smartphone Inertial Sensors. *Sensors* 2018, 18, 4132.
- [13] Meraj, S. S., Yaakob, R., Azman, A., Rum, S. N. M., & Nazri, A. S. A. (2019). Artificial Intelligence in diagnosing tuberculosis: A review. *International Journal on Advanced Science, Engineering and Information Technology*, 9(1), 81–91.
- [14] Alzubaidi, L., Zhang, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* 8, 53 (2021).

- [15] Antony, B., & Banu, N. (2017). Lung Tuberculosis Detection Using X-Ray Images. *International Journal of Applied Engineering Research*, 12, 15196–15201.
- [16] Muhathir, M., Sibarani, T. T. S., & Al-Khowarizmi, A.-K. (2020). Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (Tb) By Utilizing Hog Feature Extraction. *Al'adzkiya International of Computer Science and Information Technology (AloCSIT) Journal*, 1(1).
- [17] Dunnmon, J. A., Yi, D., Langlotz, C. P., Ré, C., Rubin, D. L., & Lungren, M. P. (2019). Assessment of Convolutional Neural Networks for Automated Classification of Chest Radiographs. *Radiology*, 290(2), 537–544. <https://doi.org/10.1148/radiol.2018181422>.
- [18] Meraj, Syeda & Azman, Azreen & Yaakob, Razali & Nazri, Azree & Rum, Siti. (2019). Detection of Pulmonary Tuberculosis Manifestation in Chest X-Rays Using Different Convolutional Neural Network (CNN) Models. 10.35940/ijeat.A2632.109119.
- [19] Gordienko, Yu., Gang, P., Hui, J., Zeng, W., Kochura, Yu., Alienin, O., Rokovyi, O., & Stirenko, S. (2018). Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. *Advances in Intelligent Systems and Computing*, 638–647.
- [20] Tasci, E. (2020). Pre-processing Effects of the Tuberculosis Chest X-Ray Images on Pre-trained CNNs: An Investigation. *Artificial Intelligence and Applied Mathematics in Engineering Problems*, 589–596.
- [21] Oloko-Oba, M., & Viriri, S. (2020). Diagnosing Tuberculosis Using Deep Convolutional Neural Network. *Lecture Notes in Computer Science*, 151–161.
- [22] Ahsan, M., Gomes, R., & Denton, A. (2019). Application of a Convolutional Neural Network using transfer learning for tuberculosis detection. 2019 IEEE International Conference on Electro Information Technology (EIT).
- [23] Colombo Filho, M. E., Mello Galliez, R., Andrade Bernardi, F., de Oliveira, L. L., Kritski, A., Koenigkam Santos, M., & Alves, D. (2020). Preliminary Results on Pulmonary Tuberculosis Detection in Chest X-Ray Using Convolutional Neural Networks. *Lecture Notes in Computer Science*, 563–576.
- [24] Showkatian, E., Salehi, M., Ghaffari, H., Reiazi, R., & Sadighi, N. (2022). Deep learning-based automatic detection of tuberculosis disease in chest X-ray images. *Polish Journal of Radiology*, 87(1), 118–124.
- [25] Jaeger, S., Candemir, S., Antani, S., Wang, Y. X. J., Lu, P.-X., & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, 4(6), 475–477.
- [26] Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning*, 448-456.
- [27] Desjardins, G., Courville, A., & Bengio, Y. (2015). Statistics of natural image categories. *Advances in Neural Information Processing Systems*, 27, 962-970.
- [28] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- [29] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on Learning Representations*.
- [30] Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1717-1724).
- [31] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*.

Automated CAD System for Early Stroke Diagnosis: Review

Izzatul Husna Azman¹, Norhashimah Mohd Saad², Abdul Rahim Abdullah³,
Rostam Affendi Hamzah⁴, Adam Samsudin⁵, Shaarmila A/P Kandaya⁶

Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malaysia^{1,3,6}

Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Malaysia^{2,4,5}

Abstract—Stroke is an important health issue that affects millions of people globally each year. Early and precise stroke diagnosis is crucial for efficient treatment and better patient outcomes. Traditional stroke detection procedures, such as manual visual evaluation of clinical data, can be time-consuming and error-prone. Computer-aided diagnostic (CAD) technologies have emerged as a viable option for early stroke diagnosis in recent years. These systems analyze medical pictures, such as magnetic resonance imaging (MRI), and identify indicators of stroke using modern algorithms and machine learning approaches. The goal of this review paper is to offer a thorough overview of the current state-of-the-art in CAD systems for early stroke detection. We give an examination of the merits and limits of this technology, as well as future research and development directions in this field. Finally, we contend that CAD systems represent a promising solution for improving the efficiency and accuracy of early stroke diagnosis, resulting in better patient outcomes and lower healthcare costs.

Keywords—Stroke diagnosis; CAD system; machine learning; deep learning

I. INTRODUCTION

Stroke, also known as cerebrovascular illness, is the third highest cause of death in Malaysia. In 2019, there were 47,928 fatalities, 443,995 common cases, and 512,726 DALYs lost due to stroke. [1] According to a White Paper on Acute Stroke Care in Malaysia by the Galen Centre for Social Health and Policy, commissioned and funded by Boehringer Ingelheim (Malaysia), only 21% of stroke patients were able to be treated at a medical facility within three hours of the onset of symptoms, while the median time from the onset of stroke symptoms to arrival at a hospital was seven hours or more [2]. Stroke is a disease that arises when an artery or blood vessel becomes clogged or ruptured, resulting in decreased blood flow and oxygen delivery to the brain which leads to temporary or permanent failure in humans; an untreated stroke can result in death [3]. Strokes are classified into two types: ischemic and hemorrhagic as shown in Fig. 1. Ischemic strokes are caused by artery blockages or occlusions caused by plaque build-up along the inner lining of the arteries. This kind of stroke occurs in more than 80% of cases, typically abruptly and without warning. Hemiparesis is the most common symptom of an ischemic stroke, which occurs when one side of the body suddenly becomes weak or unable to move. Hemorrhagic stroke is caused mostly by the rupture of cerebral blood vessels, aneurysms, or physical trauma. More than half of people who survive this type of stroke will have a serious

disability. Both the effects of a stroke and the recovery process are particular to each individual. However, only 107 neuroradiologists were supposedly available to treat patients, with at least one of them in every state hospital in our country where it is very critical and limited. Also, the traditional method of analyzing clinical data for stroke diagnosis involves manual visual inspection, which is a time-consuming process. Unfortunately, delayed stroke diagnosis and treatment can lead to brain cell death, as individuals become disabled due to a lack of oxygen and blood flow.

Technological developments in healthcare have led to various enhancements in disease diagnostics. Medical Imaging, which includes Cone Beam Computed Tomography (CBCT), Computed Tomography (CT), and Magnetic Resonance Imaging (MRI), is one of them. This technology is capable of producing specific images of a particular area based on the imaging techniques used. To diagnose stroke patients, radiologists use CT and MRI scans. However, detecting irregularities in the images can be challenging since recent research has shown that the patient's life can potentially be saved if treatment is initiated within the first six hours of a stroke. This period is referred to as the "Golden hours." [4] MRI is a medical imaging technology that employs a powerful magnet, radio waves, and computer to create highly detailed images without the use of radiation. According to Nouf Saeed Alotaibi's research in 2022, MRI is currently the most precise imaging test for brain medical imaging when compared to other technologies like CT scans or X-rays. Magnetic resonance imaging (MRI) is increasingly being used in the diagnosis and treatment of acute ischemic stroke due to its high sensitivity and relatively high specificity in detecting abnormalities that occur after such strokes [5]. The MRI process involves using T1, T2, Flair, DWI, and Black Blood sequences to gather data that must be obtained within 6 hours after the stroke that referred to as the "golden hour of stroke."

Some researchers have developed computer-aided diagnosis (CAD) systems for MRI to address the limitations of manual visual inspection for stroke diagnosis. These systems analyze MRI images using algorithms and machine learning approaches to detect anomalies associated with acute ischemic stroke. CAD systems can save time and minimize the strain on neuroradiologists by automating the analysis procedure. Furthermore, CAD systems may improve the accuracy of stroke diagnosis, resulting in earlier treatment and better patient outcomes. CAD systems for MRI stroke diagnosis are often divided into phases. MRI images are first pre-processed

to reduce noise and artefacts. Following that, feature extraction techniques are used to identify key image characteristics associated with strokes. Finally, a machine learning system is trained on a collection of MRI pictures with established stroke diagnoses to identify fresh images as normal or suggestive of acute ischemic stroke. Recent research has demonstrated that CAD systems can detect acute ischemic stroke in MRI images with great accuracy. In one study, for example, a CAD system was shown to have a sensitivity of 93% and a specificity of 95% in diagnosing acute ischemic stroke in brain MRI data. Another study discovered that a CAD system could distinguish between acute and chronic ischemic strokes with 95% accuracy. Overall, CAD systems for MRI stroke diagnosis have the potential to increase the efficiency and accuracy of stroke diagnosis, thereby shortening treatment time.

In conclusion, in the past few years, CAD systems have demonstrated great potential in enhancing the speed and precision of stroke diagnosis as well as enabling prompt and efficient treatment. This review paper aims to present a thorough summary of the latest diagnosis methods and system algorithms in CAD systems, with the goal of achieving early stroke detection. The purpose of this paper is to examine the application of these systems in modern stroke treatment research and showcase the latest developments in this area. Furthermore, this review will delve into the advantages and disadvantages of CAD systems and investigate the obstacles and prospects for additional research and enhancement. And lastly, our hope is to contribute to global efforts to enhance stroke diagnosis and treatment by conducting a critical assessment of the latest advancements in CAD systems for the early detection of stroke.

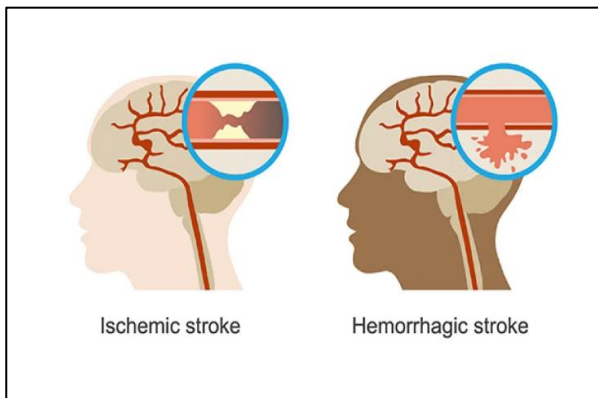


Fig. 1. Comparison of Ischemic and Hemorrhagic stroke [4].

II. MEDICAL IMAGING MODALITY FOR STROKE DIAGNOSIS

Modern advancements have resulted in a variety of methods for treating and diagnosing early stroke. These include a variety of approaches for diagnosis and assessment that doctors can use to provide the best therapy possible depending on the patient's medical history. The first step in diagnosing a stroke patient is determining if the patient is suffering from an ischemic or hemorrhagic stroke so that appropriate therapy may begin. The initial test is a CT scan or an MRI of the head for the early diagnosis of stroke.

A. Magnetic Resonance Imaging (MRI)

MRI has been the primary neuroimaging modality in several specialized clinics for critical-stage stroke in recent years, allowing pathophysiological evaluation of critical-stage stroke. In the brain, MRI can differentiate between white matter and grey matter and can also be used to diagnose aneurysms and tumours [3] [5]. MRI is the preferred imaging modality for diagnosis and therapy in the brain. Due to its sensitivity, accuracy, extension, and age, MRI is the most instructive imaging technique. Besides DWI and SWI, one of the sequences that latest in MRI is BB sequence (Black Blood Sequence) compared to conventional sequence: T1, T2 and Flair. Apart from DWI and SWI, one of the most recent MRI sequences for stroke is the BB sequence (Black Blood Sequence), as opposed to the conventional sequences: T1, T2, and Flair. Fig. 2 shows a comparison of brain lesions in BB sequence with CE T1, where study shows that BB-based automatic detection is more sensitive than MP-RAGE of CE T1 and has greater potential benefits in terms of automated CNN-based detection. MRI black blood imaging (BB) suppresses intraluminal blood signal throughout the field of view without reducing the signal strength of tiny abnormal region or lesion, allowing human readers to detect brain lesions [6]. Fig. 2 shows the comparison of the same lesion in different modalities; BB sequence and CE-T1 sequence of MRI.

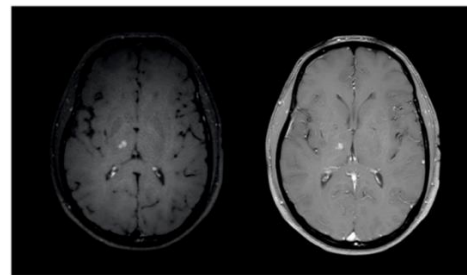


Fig. 2. Left, lesion in the Black Blood sequence (BB). Right, corresponding axial conventional Contrast-Enhance T1 slice with identical lesion. [6].

B. Computed Tomography (CT)

A CT scan is a diagnostic imaging process that produces pictures of the interior of the body using X-rays and computer technologies. It displays comprehensive views of every bodily component, including the bones, muscles, fat, organs, and blood vessels. The X-ray beam in CT moves in a circle around the body. This allows for many views of the same organ or structure and gives far more information. Hence, CT is the most effective imaging modality for acute ischemic stroke because of its availability, low cost, and quick acquisition time [7, 8]. CT also act as a standard in the early examination of acute stroke patients because it can readily and quickly see cerebral hemorrhages (as shown in Fig. 3) [9]. Dynamic contrast-enhanced (DCE)-CT is also utilized in clinical practice to improve image resolution and to examine the microvasculature of brain structures and other organs. In DCE-CT imaging, an iodine-based contrast agent is administered intravenously into the patient's body, affecting the measured X-ray absorption in tissues, and leading in contrast enhancement in the produced picture. At last in this section, Fig. 4 illustrated the three different samples of modalities including CT, Flair and DWI of MRI images.

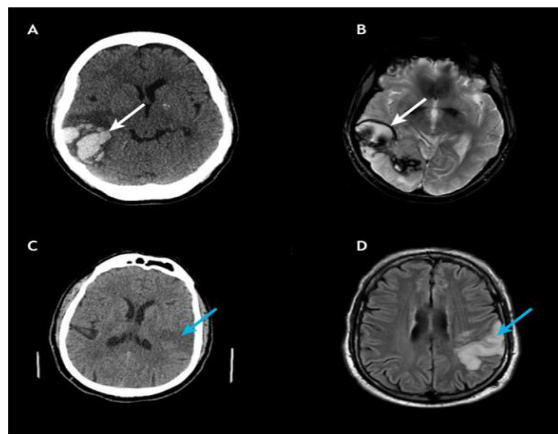


Fig. 3. CT scans against MRI scans for stroke detection. (A, B) A hemorrhage may be plainly observed in a patient's CT scan (A; white arrow), although it is less visible in an MRI scan (B; white arrow). (C, D) An ischemic infarct is just faintly visible in a C. [9].

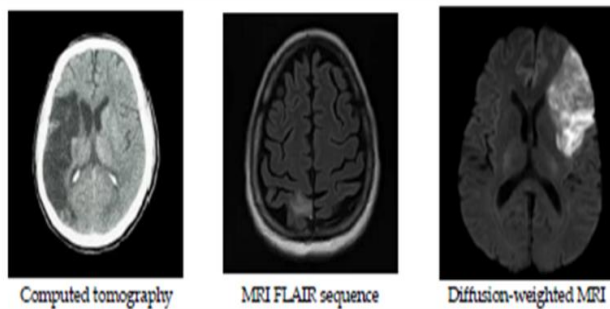


Fig. 4. Sample of CT images, MRI Flair and DWI sequences images [10].

III. STROKE DIAGNOSIS SYSTEM

Modern stroke detection devices, such as computed tomography (CT) and magnetic resonance imaging (MRI), may aid an experienced radiologist in determining if a patient has suffered a stroke; however, if a general radiologist makes an incorrect diagnosis, the patient may miss the best time for treatment. As a result, improving the quality of the diagnostic picture is crucial for assisting the physician in making an accurate diagnosis or image identification. In terms of stroke risks, different computer-aided diagnostic (CAD) systems have been created to help physicians diagnose and treat stroke patients. These methods enabled the identification of early cerebrovascular accident (CVA) symptoms and contributed to increased diagnosis accuracy for acute strokes. A notable trend is machine learning research and end-to-end system design for computer-aided stroke diagnosis and classification. Hence, there are already some important papers and developed systems for the binary classification of stroke presence or absence and hemorrhagic or ischemic stroke [11].

A. Computer-aided Diagnosis System (CAD)

The CAD technique has been used in medical imaging for illness diagnosis, prognosis, treatment decision assistance, and therapeutic monitoring. Manual segmentation in MRI takes a long time because specialists must examine several pictures of the brain taken in different orientations using different pulse

sequences [10]. Moreover, there is the possibility of inter- and intra-observer biases. These limitations can be overcome by semi-automated and automated machine learning-based CAD systems for identifying and segmenting ischemic stroke lesions, enabling high-throughput image screening for faster, reproducible, and more sensitive detection of ischemic stroke lesions. The automated delineation of the precise topology of stroke lesions allows for quantitative evaluations of infarct size and/or salvage ability, which is important for prognosis and therapeutic decision-making. A typical stroke CAD system consists of distinct sequential stages as referred to in Fig. 5.

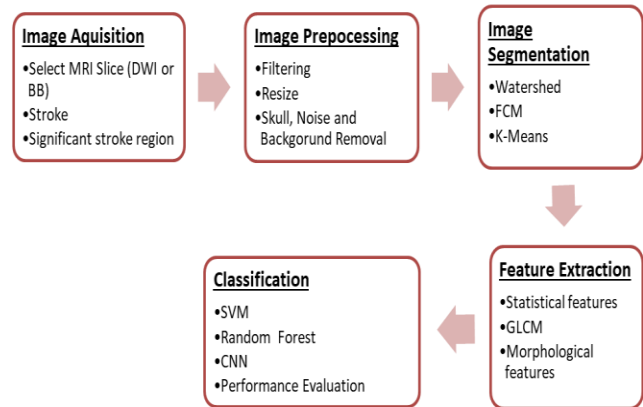


Fig. 5. Overview of typical Computer-Aided Diagnosis (CAD) system for end-to-end stroke detection.

Apart from that, several methodologies have been used to assess the performance of CAD systems in preparation for commercialization. Methods such as leave-one-out, cross-validation, hold-out, and resubstituting are examples of these techniques. The pattern recognition process in classical and current CAD medical algorithms typically consists of three major phases; however, completion of all three steps is not required. These processes are as follows: pre-processing of medical pictures, including segmentation and identification of areas of interest (ROI), extracting automatically produced or hand-engineered features from human specialists, and data categorization according to those characteristics in Fig. 5. Current CAD algorithms can offer output data without going through all of these processes; this became feasible with the development of neural networks with several hidden layers. [12]. There are also several types of processing phase in CAD as shown in Fig. 6.

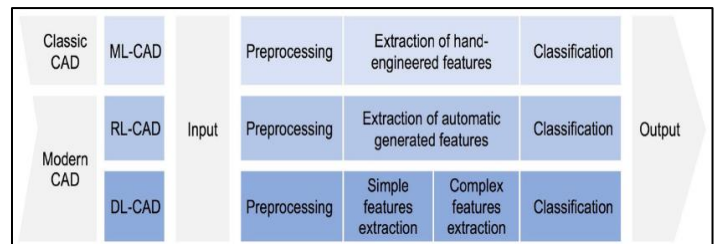


Fig. 6. Several types of processing phases in CAD including ML-CAD(Machine Learning based), RL-CAD(Representation Learning based) and DL-CAD(Dep Learning based) [12].

B. Machine Learning Technique

Watershed segmentation is a segmentation technique that uses gradients. The watershed transformation considers the size of the image's gradient as a topographic surface. The water flow line, which indicates the region's borders, corresponds to pixels with the maximum gradient magnitude intensity (GMI). Water is placed on any pixel that is surrounded by flowing stream water that flows naturally down to the minimal local average intensity. Pixels flow into the same shape as the segment's catchment basin. The watershed transform has a severe disadvantage in that it has a propensity to over-segment the picture. To evade this constraint, the object's position must be inferred using instance markers, which can guide the selection of a subset of these basins. [13]

Fuzzy C-Means (FCM) is an excellent data analysis, pattern recognition, picture segmentation, and fuzzy modelling algorithm [13]. It is a well-known soft clustering method and one of the most promising fuzzy clustering methods. Fuzzy clustering algorithms are preferable because they can more intuitively describe the link between input pattern data and clusters [14]. It is more flexible than the analogous hard-clustering approach in most circumstances, as it uses a membership function to connect each pattern to each cluster. Clustering, but not partitioning, is the result of such algorithms. A hybrid technique involving the K-means and FCM algorithms was proposed, which enhanced the accuracy in diagnosing brain infarct with lower processing costs (Fig. 7 to 9 [10]).

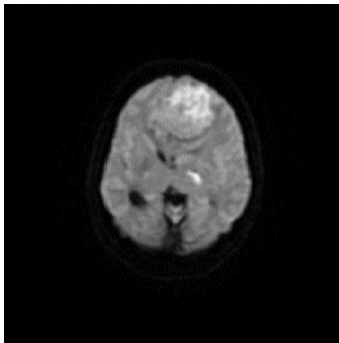


Fig. 7. Original DWI sample image.

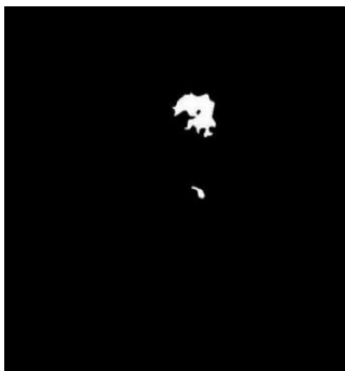


Fig. 8. Morphological binary image from sample image in Fig. 7.

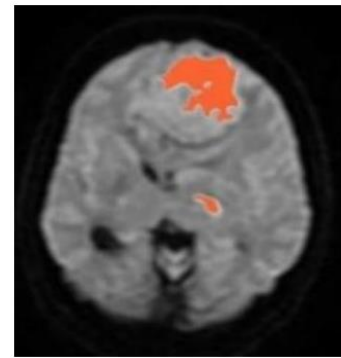


Fig. 9. Orange marking on detected infarct.

The K-Means algorithm is a clustering approach that is used to identify the optimal answer by continually decreasing the distance of components from their cluster centre and increasing the cluster. Since the k-performance algorithm is reliant on the beginning value of the cluster centre, strategies must be tested numerous times for varied results with different initial cluster centres [15]. The primary drawbacks of this approach are that the number of clusters is unexpected and that it is sensitive to the initial cluster centres. To circumvent this constraint, a statistical calculation or cluster verification approach is necessary [13]. Hence, the overview comparison in pros and cons of these 3 techniques are shown in Table I.

TABLE I. OVERVIEW COMPARISON SEGMENTATION TECHNIQUES (FCM, K-MEANS, WATERSHED) [10] [14] [15]

Segmentation Technique	Advantages	Disadvantages
Watershed Transform [10]	Several areas are segmented at the same time. It produces a complete contour of the images and avoids the need for many contours to be joined.	Over segmentation could be easily obtained.
Fuzzy C-Means [14]	Defines a distinct border for the split region.	Sensitive to noise.
k-Means [15]	If k is kept small, it is faster to compute than hierarchical clustering.	The exact number of clusters is unknown.

As for image classification in medical imaging, especially for stroke, Support Vector Machines (SVM) are classified as a classification technique, whereas it can be used in both classification and regression situations. It is capable of handling a large number of continuous and categorical variables. SVM is also classified as one of the supervised learning classification techniques that builds a hyperplane (or series of hyperplanes) in a higher-dimensional space to classify. SVM iteratively develops ideal hyperplanes, which are used to minimise errors. SVM's core notion is to determine the maximum marginal hyperplane (MMH) that optimally allocates a dataset into classes, as shown in Fig. 10. SVM was utilised to appropriately identify them as benign tumours, malignant tumours, or healthy brains. Support Vector machines (SVMs) have substantial computational advantages [16].

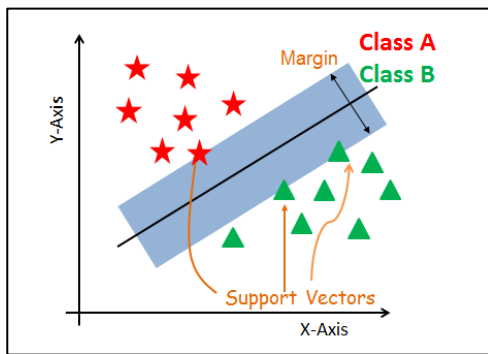


Fig. 10. Support Vector Machines techniques (SVM) [16].

Random-forest classification is a prominent supervised classification approach. Random forest classifiers are trained classifiers that are used to outline the borders of feature space points that belong to various classes, with the purpose of minimizing the margin between classes (like a SVM classifier) [17]. The random forest is made up of a few decision trees that have been trained to accept input values (feature vectors) and assign membership values to the lesion or background groups as illustrated in Fig. 11. The random forest's final predictions are formed by averaging the results of each individual tree, which tends to over fit the training data, but the random forest may reduce that issue by averaging the prediction results from multiple trees. As a result, random forests outperform single decision trees in terms of predicting validity. A study recommended that dense conditional random fields (CRF), which are frequently used as a post processing step to generate more spatially contiguous segmentations, be implemented as an optimizer in system [18]. The features used for training in RF-based multilayer cascaded RFs are still based on independent voxels and their neighbours, whereas in dense CRFs, the inference process implicitly assumes conditional independence between voxels, which may lose correlation constraints between directly adjacent voxels.

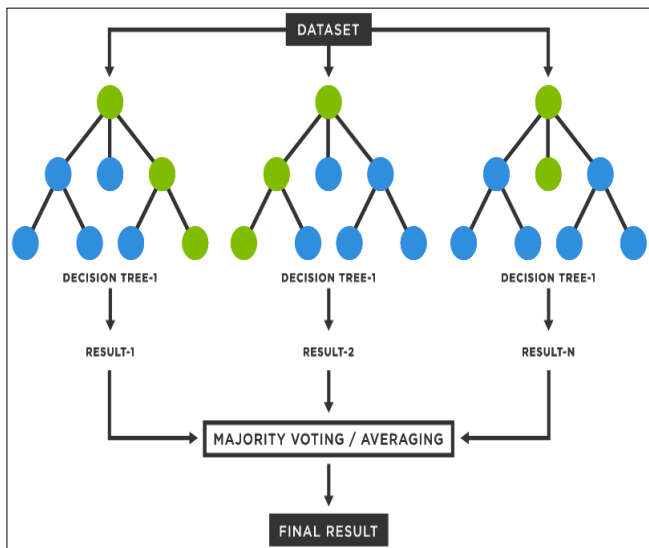


Fig. 11. Random Forest techniques (RF) [19].

C. Deep Learning Technique

Deep learning algorithms have consistently demonstrated transformative performance in a variety of tasks, most notably medical image analysis as shown in Fig. 12. Convolutional neural networks have been particularly successful in identifying and classifying patterns in medical images, leading to improved diagnosis and treatment planning. Additionally, deep learning models can also assist in predicting patient outcomes and detecting anomalies that may be missed by human observers. Convolutional neural networks are now a deep learning technique commonly used in computer vision tasks like radiography. It is designed to learn spatial information by backpropagation, utilizing numerous building blocks to achieve exceptional results in image identification. Deep learning models along with the application of CNN are being considered as methods for imaging acute ischemic strokes. Illustrated by a convolutional neural network (CNN), systematically pulls measurements from many samples in order to acquire more sophisticated abstract features for classification, recognition, and segmentation, enabling intelligent MRI interpretation [20]. It computed neural network characteristics and developed code for use in CNN input stages and passing signals with geometric detail. The visual cortex field interacts with individual cortical neurons belonging to the scanning field but does not transmit CNN weak signals. CNN nodes are linked, but not all-to-all links in a geometrical framework. In the image processing input layer, nodes are assigned to generate spectrum ranges for scanning components and shapes. All comparable structures and images (the kernel) are derived. The picture was transmitted to MRI outputs through kernel signals [7].

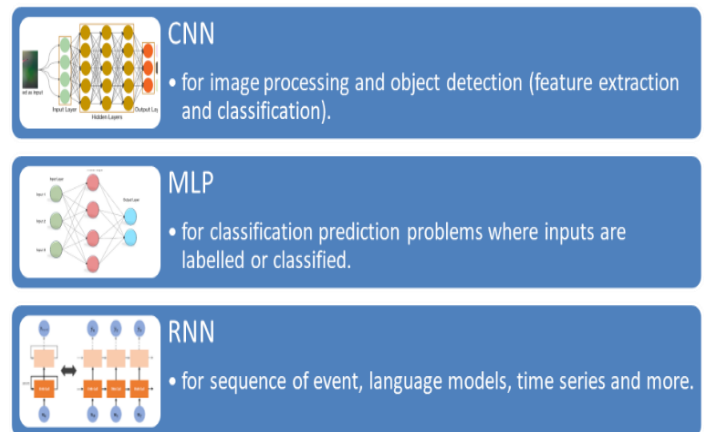


Fig. 12. Deep learning algorithm.

IV. RESULTS AND DISCUSSION

This section provides an inventory of findings from a variety of recent research initiatives that have used ML and DL approaches to construct early stroke diagnostic systems based on MRI data. We shed light on the findings of these research by thorough classification and theme analysis, uncovering recurring trends, diverse outcomes, and methodological variances as shown in Table II. The given results included a wide range of performance parameters, highlighting accomplishments in sensitivity, specificity, and diagnostic accuracy. Notably, we investigate how dataset sizes,

architectural choices, feature extraction approaches, and pre-processing procedures affect the effectiveness of these systems. We provide a panoramic picture of the state-of-the-art in ML and DL-powered stroke detection by aggregating these findings, revealing the shared discoveries that contribute to the progress of this critical medical application.

Drawing on the findings of Yannan Yu and colleagues' work, the core approach used focuses on the use of deep learning U-Net for determining Final Ischemic Stroke patients [21]. The findings of this study demonstrate a dichotomy of results, defined into two separate categories: minor instances (n = 32), with a dice score coefficient (DSC) of 0.58 (0.31-0.67), and major instances (n = 67), with a DSC of 0.67 (0.29-0.65). Furthermore, Hyuna Lee and colleagues conducted remarkable work on stroke detection, utilizing a variety of methodologies such as Logistic Regression, Support Vector Machine (SVM), and Random Forest. The findings of this comprehensive method show that Logistic Regression has a sensitivity parameter of 75.8%, while SVM and Random Forest both have a noteworthy 72.8% sensitivity [22].

Yusuf and colleagues also made significant contributions, exploring hemorrhagic and ischemic stroke using deep learning CNN approaches, especially improved MobileNetV2 and EfficientNet-B0 models. In stroke detection, the updated MobileNetV2 model obtains a remarkable accuracy of 96% (: 0.92), whereas the modified EfficientNet-B0 model achieves 93% (: 0.86). Not content with diagnosis, they expand their investigation towards vascular territorial categorization, revealing outstanding accuracy rates of 93% (0.895) using the modified MobileNetV2 model and 87% (0.805) using the modified EfficientNet-B0 CNN model [11].

Meanwhile, Sercan and colleagues focus their work on brain tumour and ischemic and hemorrhagic stroke lesion studies, using deep learning capabilities through the CNN-D-UNet architecture. Their sophisticated technique yields precision levels of 0.99 and accuracy levels of 0.989 for brain tumours, while precision and accuracy levels for brain stroke lesions are commendably documented at 0.986 and 0.985, respectively [23].

Adriell adopts a more specialised approach, concentrating just on hemorrhagic stroke segmentation using the Mask R-CNN technique and fine-tuning it to get results with a remarkable accuracy of 99.72% 0.24 and sensitivity of 99.97% 0.06 [24]. Fathia's research focuses on ischemic stroke and employs a trinity of deep learning techniques - CNN, U-Net, and Fine Tuning. Their efforts were rewarded with an average accuracy of 99.77% and a Dice Coefficient of 55.77% [25].

Finally, a notable work released in 2023 by Hongyu Gao and colleagues explores the landscape of Acute Ischemic stroke using a repertory of CNN, SVM, and Random Forest approaches. Significantly, the results highlight the CNN classifier's dominance, with an exceptional area under the curve (AUC) of 0.935, validating its usefulness in the domain [26].

TABLE II. OVERVIEW OF THE COMPARISON BASED ON RECENT RESEARCH

Author	Type of lesion	Technique used	Result parameter
Yannan Yu, 2020 [21]	Final Ischemic Stroke Lesions	U-Net	Minimal (n = 32): DSC=0.58(0.31-0.67) Major (n = 67): DSC= 0.48 (0.29-0.65)
Hyuna Lee, 2020 [22]	Stroke	Logistic regression SVM Random Forest	Sensitivity Logistic Regression = 75.8%, p: 0.02 SVM = 72.7%, p: 0.033 Random Forest = 75.8%, p: 0.013
Yusuf Kenan Cetinoglu 2021 [11]	Hemorrhagic and Ischemic stroke	CNN Modified MobileNetV2 EfficientNet-B0	Modified MobileNetV2: 96% (κ : 0.92) - EfficientNet-B0: 93% (κ : 0.86)
Sercan Yalçın 2022 [23]	Brain Tumor Ischemic and hemorrhagic stroke	CNN D-UNet	Brain Tumor: Precision: 0.99, Accuracy: 0.989 Brain Stroke: Precision: 0.986, Accuracy: 0.985
Adriell Gomes Marques 2022 [24]	Hemorrhagic stroke segmentation	Mask R-CNN Fine-tuning	Acc: 99.72 ± 0.24 Sen: 99.97 ± 0.06
Fathia ABOUDI, 2022 [25]	Ischemic Stroke	CNN U-Net Fine-tuning	Average precision: 99.77% Dice Coefficient: 55.77%
(Hongyu Gao 2023) [26]	Acute Ischemic Stroke (Endovascular Thrombectomy)	CNN SVM RF	CNN: CTP and PWI (0.902 vs. 0.928; p = 0.557)

V. CONCLUSION

In conclusion, this study found that the development of automated CAD systems for systematic stroke identification and measurement of stroke extent is necessary, which has significant therapeutic and diagnostic consequences. Eventually, this will result in better and more prompt stroke care, as well as lower patient morbidity and mortality. Advances in neuroimaging acquisition techniques, as well as the use of machine learning, are critical to this goal. In this review study, we conducted a thorough search for several image analysis approaches used to diagnose stroke lesions using MRI data. In this work, we looked at the most recent methods for segmenting and classifying cerebral stroke on MRI images, with an emphasis on machine learning approaches. This study found that by combining machine learning models and cognitive systems, brain infarcts may be recognized more efficiently and with greater accuracy on MRI in real-world clinical circumstances, which will aid in clinical

decision-making. Moreover, the limits of various methodologies as well as potential remedies are examined. We hope that this work will be a helpful resource for scholars in the subject as well as a source of ideas and inspiration.

ACKNOWLEDGMENT

The study is funded by Universiti Teknikal Malaysia Melaka (UTeM) through the Grant, No: PJP/2022/FTKKE/S01887. The authors are grateful for the facilities provided by Universiti Teknikal Malaysia Melaka (UTeM), Fakulti Kejuruteraan Elektrik (FKE), Fakulti Teknologi Kejuruteraan Elektrik & Elektronik (FTKKE), Advanced Digital Signal Processing (ADSP) Research Laboratory for all the supports.

REFERENCES

- [1] N. V. Kay Sin Tan, "Stroke Burden in Malaysia," *Cerebrovasc Dis Extra*, pp. 58-62, 2022.
- [2] A. Zainuddin, "CodeBlue Health is a human right: Malaysian Stroke Patients Take Average Seven Hours For Hospital Arrival," 28 December 2022. [Online]. Available: <https://codeblue.galencentre.org/2022/12/28/malaysian-stroke-patients-take-average-seven-hours-for-hospital-arrival/>.
- [3] E. M. S. T. H. R. B. A. M. N. M. M. A. A. A. T. S. A. a. F. A. G. Zeyad Ghaleb Al-Mekhlafi, "Deep Learning and Machine Learning for Early Detection of Stroke and," *Computers, Materials & Continua*, vol. 72, no. 1, p. 776, 2022.
- [4] U. R. ., A. G. ., Y. C. A. H. R. M. P. B. E. E. P. K. H. C. W. Y. C. J. C. R. A. Mahesh Anil Inamdar, "A Review on Computer Aided Diagnosis of Acute Brain Stroke," *Sensor*, pp. 1-35, 2021.
- [5] A. S. A. M. E. a. A. S. Nouf Saeed Alotaibi, "Detection of Ischemic Stroke Tissue Fate from the MRI Images Using a Deep Learning Approach," *Mobile Information Systems*, vol. 2022, pp. 1-11, 2022.
- [6] S. G. H. J. N. G. H. P. F. L. P. K. L. C. K. D. M. M. S. a. J. B. Jonathan Kottlors, "Contrast-Enhanced Black Blood MRI Sequence Is Superior to," *Diagnosis*, vol. 11, no. 1016, pp. 1-9, 2021.
- [7] B. Y. T. R. K. N. Surya.S, "A Comprehensive Method for Identification of Stroke using Deep Learning," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 7, pp. 647-652, 2021.
- [8] B. K. B. K. K. S. E. P. Sung-Hye You, "Fast MRI in Acute Ischemic Stroke: Applications of MRI Acceleration Techniques for MR-Based Comprehensive Stroke Imaging," *iMRI*, no. 25, pp. 81-92, 2021.
- [9] I. T. K. T. P. a. S. S. Pragati Kakkar, "Current approaches and advances in the imaging of stroke," *Disease models & Mechanism*, vol. 14, no. 12, 2021.
- [10] P. D. ., M. M. R.-S. T. U. R. A. a. S. S. Asit Subudhi, "Application of Machine Learning Techniques for Characterization of Ischemic Stroke with MRI Images: A Review," *Diagnosis*, vol. 12, no. 2535, 2022.
- [11] *. I. O. K. ., M. E. U. ., M. F. G. Yusuf Kenan Cetinoglu, "Detection and vascular territorial classification of stroke on diffusion-weighted MRI by deep learning," *European Journal of Radiology*, vol. 145, no. 110050, pp. 1-8, 2021.
- [12] J. P. D. P. d. S. C. H. a. S. N. Yahia Mokli, "Computer-aided imaging analysis in acute," *Neurological Research and Practice*, pp. 1-13, 2019.
- [13] N. S. M. N. A. R. A. Norhashimah Mohd Saad, "A Review on Image Segmentation Techniques for MRI Brain Stroke Lesion," *JTEC* 6145, pp. 27-34, 2021.
- [14] F. W. L. Z. a. G. L. Haoliang Su, "Fuzzy Clustering Algorithm-Segmented MRI Images in Analysis of Effects of Mental Imagery on Neurorehabilitation of Stroke Patients," *Hindawi*, pp. 1-10, 2021.
- [15] K. K. ., A. T. J. M. H. M. S. K. S. A. T. H. a. S. H. Abang Mohd Arif Anaqi Abang Isa, "Pseudo-colour with K-means Clustering Algorithm for Acute Ischemic Stroke Lesion Segmentation in Brain MRI," *Pertanika J. Sci. & Technology*, pp. 743-758, 2021.
- [16] V. R. S. S. V. S. Prof. Kavita Bathe, "Brain Tumor Detection Using Deep Learning Techniques," in 4th International Conference on Advances in Science & Technology (ICAST2021), Mumbai, India, 2021.
- [17] D. David, "Random Forest Classifier Tutorial: How to Use Tree-Based Algorithms for Machine Learning," 6 August 2020. [Online]. Available: <https://www.freecodecamp.org/news/how-to-use-the-tree-based-algorithm-for-machine-learning/>.
- [18] Q. L. F. S. I. R. Z. P. Gaoxiang Chen, "RFDCR: Automated brain lesion segmentation using cascaded random forests with dense conditional random fields," *Elsevier Neurimage*, pp. 1-14, 2020.
- [19] TIBCO, "What is a Random Forest?," 2023. [Online]. Available: <https://www.tibco.com/reference-center/what-is-a-random-forest>.
- [20] S. X. L. T. H. W. ., a. J. M. Shujun Zhang, "Stroke Lesion Detection and Analysis in MRI Images Based on Deep Learning," *Journal of Healthcare Engineering*, vol. 2021, pp. 1-9, 2021.
- [21] M. Yannan Yu, M. Yuan Xie, T. Thamm, P. Enhao Gong, M. Jiahong Ouyang, B. Charles Huang, P. Soren Christensen, M. Michael P. Marks, M. Maarten G. Lansberg, M. Gregory W. Albers and M. P. Greg Zaharchuk, "Use of Deep Learning to Predict Final Ischemic Stroke Lesions From Initial Magnetic Resonance Imaging," *JAMA Network Open*, vol. 3, no. 3, pp. 1-13, 2020.
- [22] P. E.-J. L. M. S. H. M. H.-B. L. M. J. S. L. P. S. U. K. M. J. S. K. M. N. K. P. D.-W. K. M. Hyunna Lee, "Machine Learning Approach to Identify Stroke Within 4.5 Hours," *STROKEAHA Journal*, vol. 51, no. 3, 2020.
- [23] H. V. Sercan Yalçın, "Brain stroke classification and segmentation using encoder-decoder based deep convolutional neural networks," *ELSEVIER (Computer in Biology and Medicine)*, pp. 1-14, 2022.
- [24] L. F. d. F. S. J. J. d. C. N. I. B. L. I. C. L. S. Adrieli Gomes Marques, "Automatic Segmentation of Hemorrhagic Stroke on Brain CT Images Using Convolutional Neural Networks Through Fine-Tuning," in 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Itali, 2022.
- [25] 2. ., C. D. a. T. K. Fathia ABOUDII, "Efficient U-Net CNN with data augmentation for MRI ischemic stroke brain segmentation," 2022 8th International Conference on Control, Decision and Information Technologies, vol. CoDIT'22, pp. 724-728, 2022.
- [26] Y. B. G. C. H. Y. Y. C. H. Z. J. W. Q. L. Q. Y. a. L. W. Hongyu Gao, "Identifying patients with acute ischemic stroke within a 6-h window for the treatment of endovascular thrombectomy using deep learning and perfusion imaging," *Frontiers in Medicine*, pp. 1-8, 2023.
- [27] T. E. T. K. H. S. D. H. S. H. B. Y. W. P. H.-J. L. B. W. C. & S. S. A. Yohan Jun, "Deep-learned 3D black-blood imaging using automatic labelling technique and 3D convolutional neural networks for detecting metastatic brain tumors," *Scientific Reports*, pp. 1-11, 2018.
- [28] D. R. S. S. Bhagyashree Rajendra Gaidhani, "Brain Stroke Detection Using Convolutional Neural Network and Deep Learning Models," in 2019 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), Manipal University Jaipur, 2019.
- [29] P. ., H. K. P. ., E. T. M. ., J. M. O. M. ., S. I. S. P. M. ., Wu Qiu, "Machine Learning for Detecting Early Infarction in Acute Stroke with Non-Contrast-enhanced CT," *Radiology*, vol. 294, pp. 638-644, 2020.
- [30] P. H. V. M. M. R. S. S. I. L. A. E. T. S. F. C. J. C. S. R. & I. Carlos Fernandez Lozano, "Random forest based prediction," *Scientific Reports*, vol. 11, no. 10071, pp. 1-12, 2021.
- [31] J. H. X. X. S. R. V. W. M. I. M. A. E. H. A. V. F. & T. S. a. V. I. i. Chin-Fu Liu, "Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke," *Communications Medicine*, pp. 1-18, 2021.
- [32] S. J. a. S. S. A. Subudhi, "Delineation of the ischemic stroke lesion based on watershed and relative fuzzy connectedness in brain MRI," *Med. Biol. Eng. Comput.*, vol. 56, no. 5, pp. 795-807, 2018.

The Current State of Blockchain Consensus Mechanism: Issues and Future Works

Shadab Alam

College of Computer Science & IT, Jazan University, Jazan, Saudi Arabia

Abstract—Blockchain is a decentralized ledger that serves as the foundation of Bitcoin and has found applications in various domains due to its immutable properties. It has the potential to change digital transactions drastically. It has been successfully used across multiple fields for record immutability and reliability. The consensus mechanism is the backbone of blockchain operations and validates newly generated blocks before they are added. To verify transactions in the ledger, various peer-to-peer (P2P) network validators use different consensus algorithms to solve the reliability problem in a network with unreliable nodes. The security and reliability of the inherent consensus algorithm used mainly determine blockchain security. However, consensus algorithms consume significant resources for validating new nodes. Therefore the safety and reliability of a blockchain system is based on the consensus mechanism's reliability and performance. Although various consensus mechanisms/algorithms exist, there is no unified evaluation criterion to evaluate them. Evaluating the consensus algorithm will explain system reliability and provide a mechanism for choosing the best consensus mechanism for a defined set of problems. This article comprehensively analyzes existing and recent consensus algorithms' throughput, scalability, latency, energy efficiency, and other factors such as attacks, Byzantine fault tolerance, adversary tolerance, and decentralization levels. The paper defines consensus mechanism criteria, evaluates available consensus algorithms based on them, and presents their advantages and disadvantages.

Keywords—Blockchain; consensus mechanism; consensus algorithm; data security; distributed systems; bitcoin

I. INTRODUCTION

The concept of blockchain revolves around the decentralized recording of digital transactions, eliminating the need for a central authority. These transactions are structured as blocks, which undergo encryption and validation by the majority of participating nodes before being appended to the blockchain. Initially introduced without standardized applications, the Blockchain methodology gained prominence with the advent of Bitcoin in 2008, credited to Satoshi Nakamoto [1]. Originally intended to circumvent the reliance on financial institutions, this innovation aimed to enable direct peer-to-peer transactions among participants. Bitcoin's success in achieving this objective set a precedent, demonstrating how businesses beyond the financial sector could conduct transactions without the intervention of a centralized third party. The structure comprises interconnected data blocks, each encapsulating transactions organized within branches of a Merkle tree, all cryptographically linked to the preceding block [2].

The blockchain operates as a ledger, capturing the complete transaction history in a chronological sequence due to the arrangement of blocks [3]. Among the most pivotal functions within the blockchain are verification and security, which are realized through a dedicated technique known as a consensus algorithm [4]. This algorithm is paramount in the blockchain system, primarily responsible for upholding its credibility, safety, and overall integrity. The consensus mechanism's efficacy directly influences critical aspects such as the stability, throughput, and accessibility of the blockchain system [5]. Within the network, nodes collaborate as validators of transactions, thereby upholding the integrity of the data. Including a block in the chain necessitates the consensus of the majority of nodes, confirming the accuracy of both the transactions contained within the block and the block as a cohesive entity. The foundation of this determination lies in a consensus algorithm implemented at the blockchain level, ensuring the precision of the data. Based on the level of access, blockchain networks can be categorized into two distinct types: private and public [6].

In contrast to public blockchains, which anybody may access and interact with, private blockchain can only be accessed by machines that have been allowed access. A consensus method in the blockchain can force the system's dispersed nodes to debate whether a transaction or block is valid. It allows for the eventual writing of valid data into the blockchain when the nodes have reached a consensus. In a distributed scheme, obtaining consensus between uncertain nodes has been discussed as a "Byzantine" problem in which a herd of army generals has cordoned off the city. Specifically, there is a clash between generals as some choose to attack, and others want to withdraw from the town. The town, assaulted by several generals, would collapse. Therefore, they should agree on whether to attack or retreat [7].

Similarly, the blockchain algorithm's major challenge in distributed ambiance is to achieve consensus [8][9]. Generally, the blockchain is decentralized because of a centralized node for noticing and checking every transaction. It creates a necessity to design and develop protocols or methods that specify all the transactions are legitimate. For this reason, the consensus algorithm is believed as the soul of every blockchain. In a decentralized or distributed environment, the consensus is a crucial issue that defines the mechanism to approve or refuse a block by every agreed node [10]. Once the new block is allowed by every network member, it is then attached to the blockchain [11]. As discussed, the blockchain's primary issue is how to achieve consensus between members of the network. Every algorithm has implemented a broad

spectrum of consensus algorithms with many strengths and weaknesses. The number of current consensus algorithms can create a fuss in choosing and applying them. Therefore, it is necessary to recognize various performance evaluation criteria that include every aspect of the consensus algorithm, besides the profound understanding of current algorithms' limitations for attaining consensus between peers and guaranteeing data security in the blockchain [12]. The main goal of this paper is to present criteria for evaluating the efficiency or the performance of widely known blockchain consensus algorithms and further review and evaluate the existing consensus mechanisms based on these identified parameters.

The subsequent sections of this article are organized as follows: Section II presents the pertinent background research in this domain. Section III offers a concise overview of the prevailing consensus techniques within the realm of blockchain technology. The approach taken to evaluate these consensus algorithms is expounded upon in Section IV. Section V presents a comprehensive analysis of the challenges and limitations inherent in these algorithms, accompanied by suggestions for potential avenues of further exploration. Section VI delves into the existing gaps and research challenges, while the conclusive Section VII provides a comprehensive summary encapsulating the entirety of this study.

II. RELATED WORK

The origins of consensus algorithms can be traced back to concepts of credibility and reliability in distributed algorithms, exemplified by the Byzantine General Problem. In 1999, Castro and Liskov introduced Practical Byzantine Fault Tolerance (PBFT), a novel consensus approach aimed at mitigating trust-related concerns. PBFT fosters trust among participating stakeholders and facilitates efficient data exchange while minimizing latency. Following this, the Proof of Work (PoW) concept emerged in the same year, drawing inspiration from PBFT's principles, and was proposed as a means of validating transactions within open distributed systems. Subsequently, the PoW concept laid the foundation for the operational model of Satoshi's Bitcoin cryptocurrency [1]. PoW involves solving complex puzzles, its functionality hinging on the value in relation to the targeted hash cost. When the cost is lower, a block is mined and subsequently appended to the blockchain.

While doing the literature review on the consensus algorithms, this article identified literature related to consensus and studies associated with comparing the consensus algorithm. To review the metrics and criteria, a systematic review of the consensus algorithms has been done. G. T. Nguyen and K. Kim reviewed the Blockchain consensus algorithms applied in some well-known applications at this time [13]. Bach et al. (2018) present a comparative study of algorithmic steps, scalability, methods, and security risks of popular consensus algorithms. Authors in [14] tested that none of the deterministic consensus protocols could guarantee a mechanism in a decentralized system. Still, Paxos can not only assure steadiness but also the security of the network. As per [15], there is no doubt that Paxos is demanding and challenging to implement and understand, but the modern training standard

allows us to achieve a consensus algorithm whenever required [16]. Paxos is the group of protocols for attaining consensus in the network of unreliable or defective processes [17]. Ferdous et al. (2020) analyze a wide range of consensus algorithms employing comprehensive taxonomic properties and investigate the consequences of the different problems that are still widespread in consensus algorithms. They also provided detailed literature on cryptocurrencies belonging to various class consensus algorithms [18]. Alsunaidi and Alhaidar thoroughly analyzed Blockchain technology, focusing on well-known consensus algorithms to identify the characteristics and variables affecting performance and security [19]. Panda et al. presented a thorough analysis of the distributed consensus processes in accordance with the kind of blockchain used. It also does a comparative analysis of the consensus protocols [20]. Sharma and Jain cover the different consensus methods, how they operate, and their applications. Additionally, it looked at blockchain technology, including its benefits and drawbacks [21].

Meneghetti et al. (2020) presented a comprehensive survey of the PoW techniques, attacks, and their current use in cryptocurrency consensus algorithms. They also analyzed some known attacks on these consensus algorithms and then presented them in a coordinated manner according to their core ideas [22]. The consensus algorithm can resolve common problems, such as harmonization among dispersed systems [23]. Consensus algorithms used in the blockchain can determine the legitimacy of distributed transactions in cryptocurrencies. Moreover, it is also used in authorizing the uniqueness of a front-runner of the distributed task. The consensus algorithm ensures reliability amongst state machine replicas and, later on, harmonizes them. The stack of 32 consensus algorithms is sorted into two significant types: proof-based and vote-based [13]. This study illustrates the advantages and disadvantages of all kinds and contrasts them, established on obtrusive characteristics.

Simultaneously, the limits and upcoming growth in technology are also discussed [13],[24]. Yang Xiao et al. (2020) survey provides comprehensive literature on blockchain consensus algorithms. The analysis is done concerning performance, fault tolerance, and vulnerabilities. At the same time, there is also an emphasis on their use cases. Bamkan et al. (2020) comprehensively examined the resources accessible on the consensus algorithms in light of their traits, motivations, and present difficulties [25]. This paper defines the criteria for consensus evaluation as throughput, profitability, degree of decentralization, and vulnerabilities and evaluates the existing blockchain consensus mechanisms based on these criteria [6].

Further, article [2] presents some open issues and challenges in implementing various consensus mechanisms with their virtues and drawbacks. In-depth research on blockchain technology has been done by examining its design, including a range of consensus algorithms and the options for security and data privacy within the blockchain discussed in this article [26]. A survey of the leading consensus mechanisms on blockchain solutions is done in this paper and highlights each one's properties. Additionally, it distinguishes between probabilistic and deterministic consensus procedures [27]. Some other studies also presented a brief review of

consensus algorithms, but these studies are not comprehensive, like [28] surveys highlighting the latest studies in blockchain and consensus algorithms. This paper adds theory and information that may be utilized to choose an appropriate consensus algorithm. It will aid scholars in their continued study of consensus in the context of private blockchain [29]. According to this article, the Byzantine consensus may need to be rethought in light of the blockchain environment, which also looks at prominent blockchain consensus algorithms [30]. Lashkari & Musilek [31] presented a very detailed analysis of existing blockchain consensus algorithms. Ferdous et al. [32]

surveyed the consensus algorithms being used in cryptocurrencies. Lina Ge et al. (2022) surveyed the PoS-based consensus algorithms and compared them with their advantages and disadvantages [33]. Xiong et al. [34] reviewed the widely used main consensus algorithms, the possible scenarios in which they can be suitable, and their relative disadvantages. Jain & Jat [35] survey some prominent consensus algorithms, reviews the key features and parameters, and compare the presented consensus algorithms based on these.

TABLE I. COMPARATIVE STUDY OF RELATED RESEARCH WORK

Ref	Year	Idea of Paper	Comments
[13]	2018	It reviews the Blockchain consensus algorithms applied for various applications.	None
[19]	2019	The author thoroughly analyzed Blockchain technology, focusing on well-known consensus algorithms to identify the characteristics and variables affecting performance and security.	It is recommended that one of the leading consensus algorithms for public Blockchain networks be improved by introducing a lightweight mechanism.
[20]	2019	This paper presented a thorough analysis of the distributed consensus processes. In accordance with the kind of blockchain used, it also does a comparative analysis of the consensus protocols.	None
[21]	2019	This paper covers the different consensus methods, how they operate, and their applications. Additionally, we looked at blockchain technology, including its benefits and drawbacks.	None
[25]	2020	This survey comprehensively examined the resources accessible on the consensus algorithms in light of their traits, motivations, and present difficulties.	It examined protocols' use cases while analyzing them in terms of fault tolerance, performance, and vulnerabilities.
[6]	2020	This paper defines the criteria for consensus evaluation as throughput, profitability, degree of decentralization, and vulnerabilities and evaluates the existing blockchain consensus mechanisms based on these criteria.	None
[2]	2020	It outlines several unresolved problems and difficulties in implementing various consensus processes and their advantages and disadvantages. The proposed poll would guide blockchain academics and developers as they consider and create the next consensus mechanisms.	None
[26]	2020	In-depth research on blockchain technology has been done by examining its design, which includes a range of consensus algorithms and the options for security and data privacy within the blockchain discussed in this article.	None
[27]	2020	A survey of the leading consensus mechanisms on blockchain solutions is done in this paper and highlights each one's properties. Additionally, it distinguishes between probabilistic and deterministic consensus procedures.	It aims to create a hybrid consensus algorithm relying on communication lines that are only partially synchronized and reaching an agreement on just allowing for one-hop neighbor voting.
[28]	2020	This survey highlights the latest studies in blockchain and consensus algorithms.	None
[29]	2020	This paper adds theory and information that may be utilized to choose an appropriate consensus algorithm. It will aid scholars in their continued study of consensus in the context of private blockchain.	To determine the actual performance indicators of the consensus employed, additional study can be conducted by adjusting the number of loads and peers and assessing it using some benchmarks.
[30]	2020	According to this article, the Byzantine consensus may need to be rethought in light of the blockchain environment, which also looks at prominent blockchain consensus algorithms.	None
[31]	2021	Presented a very detailed analysis of existing blockchain consensus algorithms.	It does not consider the attacks on consensus algorithms.
[32]	2021	It surveys the consensus algorithms being used in crypto-currencies.	It does not consider the attacks on consensus algorithm and consider only crypto-currencies.
[33]	2022	Survey on consensus algorithm for Proof of Stake (PoS)	Discussed only PoS-based consensus algorithm
[34]	2022	Presents the review of main consensus algorithms being widely used, the possible scenarios in which they can be suitable, and their relative disadvantages.	It does not consider the attacks on consensus algorithms.
[35]	2022	This paper surveys some prominent consensus algorithms, reviews the key features and parameters, and compares the presented consensus algorithms based on these.	A limited no of consensus algorithms are taken and further does not consider the attacks on consensus algorithms in detail.
Our Review	2023	Our paper conducted a detailed review of the maximum prominent blockchain consensus algorithm. It further compared these consensus algorithms based on performance and security attack criteria.	Other articles have either covered the security attacks or performance analysis but have not combined both approaches.

Recent surveys on consensus algorithms have examined the limitations and future work of various consensus algorithms. Nonetheless, there is a gap in the existing analysis of the consensus algorithm. The current literature does not provide enough criteria for a comprehensive and comparative analysis of consensus algorithms. Henceforth, this paper aims to provide a complete and detailed analysis of existing and recent consensus algorithms concerning throughput, scalability, latency, and energy efficiency. Table I present a comparative study of previous related and current research work and highlights the significance of the recent research work.

This paper also takes other factors, including attacks, Byzantine fault Tolerance, adversary tolerance, and decentralization levels. Besides comparison, this paper presents the advantages and disadvantages of consensus algorithms. The analysis results are shown in tabular formats, visually illustrating these algorithms in a meaningful way.

III. PARAMETER FOR EVALUATION

This sub-division will discuss different parameters that categorize consensus algorithms [2].

A. Blockchain Type

Blockchain can be categorized into three primary classes: private, public, and consortium. These classifications are indicative of the governance structure among participants and the specific nature of the blockchain.

B. Scalability and Attacks

In decentralized systems, scalability plays a vital role. In terms of scalability, consensus algorithms are separated, like ELASTICO and Proof of Trust, but PoW is non-scalable.

C. Adversary Tolerance

It quantifies the blockchain's ability to withstand malicious operations. Additionally, it gauges the stability of the blockchain network during catastrophic events. Research has demonstrated that the consensus algorithm exhibits the highest level of tolerance towards adversaries.

D. Throughput

Throughput in the consensus algorithm means how long it takes to confirm the transactions in a blockchain network [36]. It further suggests that the extreme throughput is an absolute rate at which the blockchain can authorize transactions [37].

E. Energy Consumption

Out of the various factors or criteria that disturb the blockchain consensus algorithm's valuation is power utilization. There is a variation in consensus algorithms' energy consumption that cannot be experimentally evaluated due to varied heterogeneous limitations [38].

F. 51% Attack

A 51% attack is commonly known as an assault on a blockchain, typically targeting bitcoins, executed by a group of miners wielding over 50% of the network's mining hash rate or computational power [6]. Usually, these types of threats cannot be evaded theoretically [39]. Blockchain protocols strive to elevate the costs associated with this attack to deter it, although a complete resolution remains elusive.

G. Double Spending Attack

A double-spend is a unique problem related to digital currencies that works when one user spends the digital assets more than once [40]. Since there is no centralized authority to control transactions, the attacker will attempt to generate a regular contract to contain it in a block. Then he will try to outspread the deceitful branch of the system he had shaped until the deceitful branch is confirmed and accepted as the precise branch that consists of the fraudulent transaction [41].

IV. REVIEW OF EXISTING CONSENSUS ALGORITHM

In simple language, the term consensus means harmony or concord. The consensus algorithm will authorize an agreement among all the nodes, thereby guaranteeing reliability and trust between the unidentified peers. The consensus algorithm also ensures that each block in the existing chain involves every peer node across the system [38]. That enables distinctness and clarity in the added processes or transactions, which defines a mutually beneficial network for every node. It is worth noting that once the block gets verified, it's practically impossible to eliminate or alter them. The consensus algorithm erases all the non-member intermediaries to guarantee the accuracy of the transaction [3]. However, once the consensus involving chain transactions obtains a global status, all nodes or peers become reliable for the blockchain structure. It eventually helps in the authentication of the untrustworthy and uncertain network associated with the self-contradictory person. However, in this part, we will present the utmost significant consensus algorithms commonly utilized in the blockchain system, with their disadvantages and benefit in general.

A. Proof of Work (PoW) [1]

It was presented by Nakamoto and later applied to Bitcoin [1]. Subsequently, this was endorsed by other cryptocurrencies, which include Ethereum, Dogecoin, Monero, and last but not least, Litecoin. It has a high algorithmic cost with a clear quorum design. Hash is a difficult and random mathematical formulation used to confirm the saved operation within blocks [42]. To achieve consensus in a network, miners strive against each challenging computational puzzle. Such puzzles are challenging to solve, but the result can be promptly verified once they are solved. Once the miner found the solution to the new block, it is broadcasted to the network. In turn, all other miners will confirm and verify that the solution is accurate, and then the block may be confirmed [43][2]. The PoW algorithm's benefit is that it comes with a significant amount of security, a decentralized framework, and a permissible level of scalability. On the contrary, it has some disadvantages, including lesser throughput, high block creation time, the inadequacy of energy, dependencies on specialized hardware, high computation cost, and comprehensive bandwidth [9], [19].

B. Proof of Stake (PoS) [33]

It arose as a substitute for PoW, originally used as a consensus algorithm in blockchain technology, and was applied to validate and add new blocks to the chain. PoW requires enormous amounts of energy, which is the main reason for PoS establishment. For this reason, the authors suggested light-weighted consensus protocols for lower-power IoT communication channels [44]. PoS is based on the concept

that individuals can confirm or excavate block transactions according to how many coins they retain [45]. The miners will obtain no award besides the transaction fee in these methods. If the full node is chosen to build a new block, then the lender will gain a proportion of those operations [6].

C. Distributed Proof of Stake (DPoS) [46]

It was introduced by Daniel Larimer [47]. A key feature of this algorithm is its emphasis on decentralization. DPoS structures the network more efficiently, granting each delegate ample time to publish on every node [2]. This approach finds utility in private blockchains due to its semi-centralized characteristics. Within this method, potentially malicious miners are subject to capping based on specific parameters such as intervals and block sizes [9].

D. Practical Byzantine Fault Tolerance (PBFT) [48]

PBFT deals with the byzantine issue of the distributed nodes that can cause 33% of work damage because of chain faults. PBFT is the capability of a distributed network to reach an adequate consensus despite malignant nodes in a system failure or the broadcast of incorrect information. PBFT aims to safeguard against disastrous system failure by decreasing the effect of the malignant nodes [49]. The advantage of this method is its high throughput and energy efficiency. On the other hand, specific points like scanty or no scarce constraints quantifiable for being scalable and network delays while stating every node poll are some of its disadvantages.

E. Proof of Importance (PoI) [50]

In PoI, a miner's application-specific integrated circuit chips are deployed to enhance computing power. It works when any more family of coins has a strong possibility to mine the next block. PoI compensates users with more transactions and the user with a considerable net stake in tackling these restrictions. PoI was first established in the NEM design [50]. In PoI, each node is allocated a significant value. A node carrying out a transaction with a node with great significant worth is, in all probability, to mine the next block even though the node has less stake than another node. It is considered an improvement over the PoS algorithm [2].

F. Proof of Capacity (PoC) [51]

It was introduced in 2015 by Dziembowski. As the name implies, PoC's dynamics revolve around selecting a miner node based on the available memory capacity of an external hard disk. The node with a larger storage capacity can precompute and retain a greater number of solutions for the impending problem before actual mining begins. This approach effectively addresses the intricate challenges associated with node management within the Proof of Work framework, subsequently alleviating broader difficulties. PoC entails the strategic utilization of hard drive resources, encompassing the storage and computation of results on the hard drive prior to the commencement of the mining process.

G. Proof of Burn (PoB) [52]

This method is a substitute for attaining a deal in the blockchain network. This algorithm node in the network has to lose or scorch cryptocurrency to obtain the mining entitlement to the permitted source. This method is less like Proof of Work,

but the only difference is where the belongings are in the form of cryptocurrency rather than the computing power of a node. The loss of coins reflects the node's longer commitments to stay sincere in the system as it has lost real coins to increase the mining entitlement [2].

H. Delegated Byzantine Fault Tolerance (DBFT) [2]

It can be derived that DBFT monitors the conventional phases of the DPoS protocol in the start-up phase. In this method, the consensus is obtained using a superannuated BFT method by adding extra steps [2]. Here the user will vote and select members to add the new role in the chain based on bulk voting of more or equal to 66% affirmative from the members [42]. It should be noted that fault tolerance of delegated Byzantine is very rarely prone to confront delays from the PBFT, but restricting the number of votes can jeopardize the decentralization of the network [4].

I. Reliable, Replicated, Redundant, And Fault-Tolerant (RAFT) [53]

This method is a Substitute for the Paxos protocol. This method is more straightforward and, at the same time, provides safety and privacy with add-on features[2]. The consensus in this method is reached by choosing a delegate, and then this delegate will be accountable for copying the logs every time the latest user accesses the network. Heartbeat notes will operate as an interfering signal for marking the presence of the forerunner [2]. Each node will have a time-out for the signal's arrangement if it will not get the message before its lapse. After this, there will be a process of selecting the new leader, or else time will reset.

J. Proof of Activity (PoA) [54]

One more consensus algorithm, PoA, was developed by Bentov et al. in the year 2014 [55]. The authors mentioned this algorithm as a union of PoS and PoW. It is a safer algorithm countering Bitcoin's potential assaults and has even ignorable sanctions concerning the network communication and storage area. Nevertheless, through PoS structured protocols, shareholders may engage in downward price spirals; for that reason, the coins that they maintain will produce revenue commensurate to real commerce taking place [2].

K. Proof of Authentication (PoAH) [56]

It is a consensus algorithm aimed at a lightweight and sustainable blockchain for building a lightweight decentralized security system to circumvent central dependencies. PoAH is a cryptographic verification mechanism that is a replacement for the PoW algorithm. This consensus algorithm is appropriate for private and permissible blockchain and makes blockchain application-specific. Besides securing the system, PoAH maintains sustainability and scalability.

L. Proof of PUF-Enabled Authentication (PoP) [57]

It is a comprehensive algorithm that effectively manages both data and device security aspects. This innovative approach combines the utilization of physical unclonable functions (PUF), which serve as integral hardware security components. These PUFs contribute to the system's ability to offer advantages in terms of latency, scalability, and energy consumption. The mechanism involves incorporating a

cryptographic hash of all previously processed data along with the involvement of any device incapable of generating the PUF key in a uniquely generated manner within the PUF module. This integrated approach ensures the robust handling of both data and security keys. In comparison to Proof of Work (PoW), PoP demonstrates a notable increase in speed, while in contrast to Proof of Authority and Hashpower (PoAH), it exhibits a slightly elevated latency.

M. Rock-Scissors-Paper (RSP) [58]

To achieve consensus and avoid attacks by the malicious participant, this algorithm uses three balance variables: Rock, scissors, and Paper. RSP does not directly address the problem of variable difficulty; instead, it proceeds with the consensus based on the device's specification. Furthermore, computations can be performed quickly and easily using a high specification of computing devices. This consensus algorithm reduces the power utilization that is required to limit the maintenance and processing cost.

N. Proof of Research (PoR) [18]

It is a hybrid consensus algorithm that combines Proof of Stake (PoS) with Proof of BOINC (Berkeley Open Infrastructure for Network Computing). This innovative approach is facilitated by Gridcoin, a cryptocurrency that individuals can acquire through the contribution of their computational resources to the BOINC project. PoR bears similarities to PoS, allowing individuals to become investors by possessing a designated quantity of Gridcoin and engaging in the minting process.

O. Proof of Stake Velocity (PoSV) [18]

It is an innovative consensus algorithm crafted to address the challenges encountered within the Proof of Stake framework. PoSV introduces a hybrid approach that integrates seamlessly with conventional PoS algorithms. The fundamental premise of PoSV lies in the concept of stake velocity, which mirrors the concept of money velocity in economics. The core principle driving stake velocity is the augmentation of stake circulation during the PoS consensus process. Investors can actively enhance this stake flow by engaging in the consensus mechanism, thereby staking their cryptocurrency as a dynamic alternative to passively holding it offline. This strategic involvement substantially enhances the security measures and

mitigates the issue of inadequate participant engagement often observed in conventional PoS systems.

P. Proof of Familiarity (PoF)[59]

This consensus algorithm is designed to integrate various healthcare stakeholders' medical conclusions. PoF guarantees stakeholders' medical results' privacy and integrity by utilizing previously-stored results using blockchain. Proof of familiarity uses a two-layer security measure to preserve the identity of stakeholders. It first stores stakeholders' identities locally, and then the hash of these are stored in the blockchain.

Q. Proof of Trust (PoT) [60]

Consensus protocol integrates a confidence dimension to satisfy the service sector's practical criteria, i.e., fixing the unfaithful activities that exist so frequently in a transparent, public service network, together with the reward steps. PoT consensus utilizes random logic algorithms to maximize block node unpredictability using time signs and digital signatures. A credibility evaluation of the crowdsourcing membership involved will be done automatically by the improved algorithm. The validity, equity, and stability can be obtained by the PoT.

R. Proof of Luck (POL) [61]

PoL is a blockchain consensus algorithm that uses a random number generator on a trusted execution environment (TEE) platform to select a consensus leader. This allows for fair mining while also enabling quick transaction validation, deterministic confirmation times, and low energy consumption, among other benefits.

S. Leased Proof of Stake (LPoS) [61]

It represents a variant of the PoS consensus mechanism. Notably employed within the Waves platform, this distinctive PoS approach facilitates token holders in "leasing" their tokens to complete nodes, thereby earning a share of the rewards. On conventional PoS networks, individual nodes contribute new blocks to the blockchain. Within the LPoS framework, users have the flexibility to actively operate a full node or alternatively lease their stake to a full node. This engagement in the LPoS ecosystem yields rewards for the participants.

Table II illustrates the comparison of various consensus algorithms on defined parameters.

TABLE II. ANALYSIS OF CONSENSUS ALGORITHMS

Consensus Algorithms	Byzantine fault Tolerance	Adversary tolerance	Decentralization level	Node identity	Throughput(tps)	Scalability	Latency	Energy efficiency	51% Attack	Double spending Attack	Trust
PoW [62]	50%	<25%	Decentralized	Permissionless	Low	High	High	No	Vulnerable	Vulnerable	Untrusted
PoS [33]	50%	<51%	Semi-Centralized	Permissionless	Low	High	Medium	Yes	Vulnerable	Difficult	Untrusted
DPoS[46]	50%	<51%	Semi-Centralized	Permissioned	High	High	Medium	Yes	Vulnerable	Vulnerable	Trusted
PBFT[48]	<=33%	<33%	Decentralized	Permissionless	High	Low	Low	Yes	Safe	Safe	Semi-trusted
PoI[50]	50%	N/A	Decentralized	Permissionless	Low	High	Medium	Yes	Safe	Safe	Untrusted
PoC[51]	NA	NA	Decentralized	Permissioned	Low	High	High	Fair	Vulnerable	Vulnerable	Semi-trusted
PoB[52]	NA	<25%	Decentralized	NA	Low	Low	High	No	Vulnerable	Vulnerable	Untrusted
DBFT[2]	NA	<33%	Semi-Centralized	Permissionless	High	High	Medium	Yes	Vulnerable	Vulnerable	Semi-trusted
RAFT[53]	>50%	<50%	Decentralized	Permissionless	High	High	Low	Yes	NA	Safe	Trusted
PoA[54]	>50%	N/A	Decentralized	Permissionless	High	High	Low	No	Vulnerable	Vulnerable	Trusted
PoAh[56]	N/A	N/A	Decentralized	Permission-based	N/A	High	low	low	No known attacks	No known attacks	Trusted
PoP[57]	N/A	N/A	Decentralized	Permissioned	N/A	high	low	low	No known attacks	No known attacks	Trusted
PoR[18]	50%	<51%	Semi-Centralized	Permissionless	Medium	low	medium	medium	Vulnerable	difficult	Untrusted
PoSV[18]	50%	<51%	Semi-Centralized	Permissionless	high	medium	low	low	Vulnerable	difficult	Untrusted
RPS[58]	N/A	N/A	Decentralized	Permissioned	N/A	N/A	N/A	low	No known attacks	No known attacks	Trusted
PoF[59]	N/A	75%	Decentralized	Permissioned	medium	high		low	No known attacks	No known attacks	Trusted
PoT[60]	>50%	N/A	Decentralized	Permissioned	high	high	low	medium	safe	safe	trusted
PoL[61]	N/A	<25%	Decentralized	N/A	high	high	low	yes	safe	safe	trusted
LPoS[61]	N/A	<51%	Decentralized	Permissioned	high	high	high	yes	No known attacks	No known attacks	Semi-trusted

V. ANALYSIS OF CONSENSUS ALGORITHM

The foundation of blockchain rests on a secure and dependable architecture that stems from consensus mechanisms. Different consensus algorithms are applied to specific applications due to the unique demands of each domain. For instance, some domains require swift transaction processing, while others prioritize minimal computational power consumption. The consensus algorithm assumes a pivotal role within the blockchain framework. It operates on the premise that consensus is crucial to achieving unanimous agreement among network nodes during the process of block authentication [63]. The consensus algorithm strives to strike a balance among miners, assigning them equal weight to facilitate arriving at a resolution or decision by the majority of miners.

However, while this approach suits controlled environments, it proves inadequate for public blockchains as it exposes vulnerabilities to Sybil attacks. These attacks involve an individual creating multiple identities to manipulate the blockchain's functioning. In a decentralized ecosystem, a single block's addition is the responsibility of a single participant. The user selection process can be either random or based on specific criteria. Nevertheless, relying on random selection leaves the system susceptible to potential breaches.

Since blockchain is a decentralized network, no single node can handle the entire network. That is why blockchain has

endorsed a distributed consensus method to implement the data's uniformity and trustworthiness [64]. PoW [62] is based on the idea that nodes are less likely to attack the network as long as they invest a significant amount of computational effort. In a PoW blockchain, miners must perform computationally intensive tasks to add a block, making it nearly impossible for Sybil attacks to occur. PoW operates through a process called mining, where nodes perform calculations until the correct result is found. In the case of Bitcoin, the mining process involves searching for a random number, or nonce, that generates the correct hash for a block header. Therefore, the miners should be able to carry out specific tasks to calculate the figure. Once the miner overcomes the issue, all the other nodes are responsible for confirming that the response is accurate. Because of the more utilization of energy in PoW, its rendering becomes ineffective in the lower-powered application. Moreover, the nodes that take part in the block's authentication shall not correspond to enhancing the transactions of a block that makes PoW non-scalable [65].

PoS-Proof of Stake creates division among its users based on stake [33]. Any node with a definite volume of stake in their blockchain could be the miner. This algorithm also reassumes that any extra stake user will be less susceptible to a network attack. When any node turns out to be a miner, it will assign a particular quantity of its stack; therefore, a network holds this volume to ensure the user is trustworthy and permissible to do the mining. PoS needs significantly less computing energy, so

Proof of Stake has a very low power utilization than PoW. The only problem with PoS is that the mining procedure always aims at its richest member because they own a more significant stake over the rest of the nodes. DPoS, or Delegated Proof of Stake, is an additional consensus method projected to improve PoS [46]. In this method, only limited members are accountable for validating the blocks rather than only transferring this responsibility to stakeholders. The main advantage of DPoS is quick transactions because fewer nodes participate. Moreover, the selected nodes are capable enough to fine-tune the size of the block and the intervals. Fraudulence can be dealt fastly since substituted nodes are replaced with ease. One more type or alternative of PoS is TaPoS (Transaction as Proof of stake) [66]. Contrary to PoS, in which only a limited number of nodes can assist the security of a network, in TaPoS, each node secures the network. The disadvantage of PoS is the accumulated stack age, although the node is not linked to the network. PoA is projected to compensate nodes based on what activity does and their network ownership [67].

PBFT is projected to aim at asynchronous situations to help in solving BG (byzantine general) problems [48]. This method presumes that beyond two-thirds, all nodes are genuine, and beneath this are malevolent. A front runner gets selected by every block of the family, and then this front runner or leader's job is to validate a block. Another alternative to the BFT is Delegated BFT (DBFT), which works like a DPoS in which only a few nodes are accountable for authenticating and generating the block. One more protocol that is quite similar to PBFT is SCP (Stellar Consensus Protocol). SCP is carried out based on a method or algorithm named FBA (Federated Byzantine Agreement) [68]. The only alteration between PBFT and PCA is that PBFT entails a contract from widely held nodes, whereas SCP depends on a subsection of the nodes, which are considered very important. Table III below presents a concise comparison based on their respective advantages and disadvantages.

TABLE III. COMPARISON OF CONSENSUS ALGORITHMS BASED ON THEIR ADVANTAGES AND CHALLENGES

Consensus Algorithms	Advantages	Challenges
PoW [62]	*Extensive power of decentralization *Extra protected network	*High drafting power (expensive) * High electricity utilization
PoS[66]	*Energy efficient & faster processing * Improved rewards & more significant stakes	*Less decentralization than PoW *Less security than PoW
DPoS[46]	*accelerated processing than PoW and PoS * Enhanced recompenses allocation and energy efficient	*More prone to attacks and is less decentralized *Affluent people control the network
PBFT[48]	*Capable of doing transactions devoid of confirmation *Substantially reduce energy	*Elevated volume of connection between nodes *Difficult in the message's authenticity and is prone to Sybil assaults.
PoI[50]	*Quick and power-efficient *no particular hardware is required for mining.	N/A
PoC[51]	*larger drive sizes	N/A
PoB[52]	*PoB enforcement can be tailored *The power of burnt coins diminishes fractionally every time a fresh block is mined	*Source waste (the burnt coins are lost) *Huge risk protocol, no coin retrieval assurance
DBFT[2]	*Provides perfect decisiveness *Quick transaction delivery	*Prone to 51% attack *Still believed centralized
RAFT[53]	*Could endure catastrophe of up to half of the nodes *Structure clarity and robustness	*Present execution can ensure liveness for one Byzantine failure
PoA[54]	*High security & low transaction fee *Eliminates 51% attack in the blockchain network	*Requires a significant number of assets in the mining phase *Participants can double-sign transactions
PoAh[56]	*Appropriate for private as well as permissioned blockchain *Maintains system sustainability and scalability	N/A
PoP[57]	*PoP is highly scalable *Runs noticeably faster, consumes fewer resources and uses less energy.	N/A
PoF[59]	*The integrity of a medical conclusion.*Privacy of participants	N/A
PoSV[18]	*Raise the overall security of the system *Counter the lack of participant issues in PoS	*Less decentralization than PoW
PoR[18]	*Faster and Energy efficient	*Less security than PoW
RPS[58]	*Efficient power consumption and economical maintenance cost.*Fast processing time	*Specification of devices can result in the polarization of the computing devices.
PoT[60]	*highly scalable*ensures the performance and consistency of the consensus process.	N/A
PoL[61]	*Extensive power of decentralization.*low-latency transaction validation.	*Attacker may confront a limited number of TEEs
LPoS[61]	*Energy efficient & faster processing *Improved rewards & more significant stakes	*Less decentralization than PoW *Less security than PoW

VI. OPEN ISSUES AND RESEARCH CHALLENGES

Some of the open issues and research challenges are emphasized in this section.

A. Overhead

Blockchain introduces significant overhead in terms of traffic, encompassing factors such as storage size, heightened implementation costs, legal compliance considerations, and deficits in information and organization. It presents a substantial challenge, particularly with regard to escalating energy consumption.

B. Cross-compliant Hybrid Alternative (CHA)

Although many providers favor creating consensus solutions based on particular use case requirements, consensus mechanisms still need to handle various requirements. As a result, the CHA class is anticipated to witness a large variety of consensus mechanisms [31].

C. Hybrid Consensus Algorithms

A single particular type of consensus algorithm frequently has more restrictions in practical application scenarios. Examples include the PoW algorithm's resource consumption issue and the PBFT algorithm's difficulty applying only to consortium and private chains, not public ones. The goal of maximizing strengths and avoiding weaknesses can thus surely be achieved by combining the advantages of multiple algorithms into one. Additionally, this offers a fresh concept and point of reference for advancing consensus algorithms in the future [34].

VII. CONCLUSION

Recent surveys on consensus mechanisms have analyzed the performance and application set-ups, limitations, and future work of various consensus algorithms. Nonetheless, there is a gap in the existing analysis of the consensus algorithm. This paper provides a complete and detailed analysis of current and recent consensus algorithms based on throughput, scalability, latency, and energy efficiency. Further, this paper also evaluates the consensus algorithms based on 51% attacks, Byzantine fault Tolerance, adversary tolerance, and decentralization levels. Besides comparison, this paper presents the advantages and disadvantages of consensus algorithms to understand existing research challenges clearly. This comparison also highlighted the resource requirements for choosing a suitable consensus algorithm for a resource constraint environment. The analysis results have been presented in tabular formats, visually illustrating these algorithms in a meaningful way. These evaluations reflect that PoAh, PoP, PoT, and PoI are promising approaches that have high Byzantine fault Tolerance, and no known attack has been reported till now against these consensus mechanisms. This article has further highlighted the open issues and research challenges affecting the consensus mechanism. These open issues and research challenges can be further researched in detail for future research.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

- [2] U. Bodkhe, D. Mehta, S. Tanwar, P. Bhattacharya, P. K. Singh, and W. C. Hong, "A survey on decentralized consensus mechanisms for cyber physical systems," *IEEE Access*, vol. 8, pp. 54371–54401, 2020, doi: 10.1109/ACCESS.2020.2981415.
- [3] S. Alam et al., "Blockchain-based Initiatives: Current state and challenges," *Comput. Networks*, vol. 198, p. 108395, 2021.
- [4] T. Aslam et al., "Blockchain based enhanced ERP transaction integrity architecture and PoET consensus," *Comput. Mater. Contin.*, vol. 70, no. 1, pp. 1089–1109, 2022, doi: 10.32604/cmc.2022.019416.
- [5] H. Qin, Y. Cheng, X. Ma, F. Li, and J. Abawajy, "Weighted Byzantine Fault Tolerance Consensus Algorithm for Enhancing Consortium Blockchain Efficiency and Security," *J. King Saud Univ. Inf. Sci.*, 2022.
- [6] S. M. H. Bamakan, A. Motavali, and A. Babaei Bondarti, "A survey of blockchain consensus algorithms performance evaluation criteria," *Expert Systems with Applications*, vol. 154, p. 113385, Sep. 2020. doi: 10.1016/j.eswa.2020.113385.
- [7] S. Alam, "Security Concerns in Smart Agriculture and Blockchain-based Solution," in *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*, 2023, pp. 1–6.
- [8] Y. Liu, Z. Zhao, G. Guo, X. Wang, Z. Tan, and S. Wang, "An identity management system based on blockchain," *Proc. - 2017 15th Annu. Conf. Privacy, Secur. Trust. PST 2017*, pp. 44–53, 2018, doi: 10.1109/PST.2017.00016.
- [9] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends," in *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, Jun. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.
- [10] M. Shuaib, N. H. Hassan, S. Usman, S. Alam, N. A. A. Bakar, and N. Maarop, "Performance Evaluation of DLT systems based on Hyper ledger Fabric," *2022 4th Int. Conf. Smart Sensors Appl.*, pp. 70–75, Jul. 2022, doi: 10.1109/ICSSA54161.2022.9870957.
- [11] M. Shuaib et al., "Land registry framework based on self-sovereign identity (SSI) for environmental sustainability," *Sustainability*, vol. 14, no. 9, p. 5400, 2022.
- [12] M. K. I. Rahmani et al., "Blockchain-based trust management framework for cloud computing-based internet of medical things (IoMT): a systematic review," *Comput. Intell. Neurosci.*, vol. 2022, 2022.
- [13] G. T. Nguyen and K. Kim, "A survey about consensus algorithms used in Blockchain," *J. Inf. Process. Syst.*, vol. 14, no. 1, pp. 101–128, 2018, doi: 10.3745/JIPS.01.0024.
- [14] M. J. Fischer, N. A. Lynch, and M. S. Paterson, "Impossibility of Distributed Consensus with One Faulty Process," *J. ACM*, vol. 32, no. 2, pp. 374–382, Apr. 1985, doi: 10.1145/3149.214121.
- [15] L. Lamport, "The part-time parliament," in *Concurrency: the Works of Leslie Lamport*, Association for Computing Machinery, 2019. doi: 10.1145/3335772.3335939.
- [16] D. Ongaro and J. Ousterhout, "In search of an understandable consensus algorithm," in *Proceedings of the 2014 USENIX Annual Technical Conference, USENIX ATC 2014*, 2019, pp. 305–319.
- [17] B. Turner, "The Paxos Family of Consensus Protocols," 2007, [Online]. Available: <http://www.fractalscape.org/files/paxos-family.pdf>
- [18] M. S. Ferdous, M. J. M. Chowdhury, M. A. Hoque, and A. Colman, "Blockchain Consensus Algorithms: A Survey," pp. 1–39, 2020.
- [19] S. J. Alsunaidi and F. A. Alhaidari, "A Survey of Consensus Algorithms for Blockchain Technology," in *2019 International Conference on Computer and Information Sciences (ICIS)*, Apr. 2019, pp. 1–6. doi: 10.1109/ICCISci.2019.8716424.
- [20] S. S. Panda, B. K. Mohanta, U. Satapathy, D. Jena, D. Gountia, and T. K. Patra, "Study of Blockchain Based Decentralized Consensus Algorithms," vol. 2019-Octob. *IEEE*, 2019, pp. 908–913. doi: 10.1109/TENCON.2019.8929439.
- [21] K. Sharma and D. Jain, "Consensus Algorithms in Blockchain Technology: A Survey," *IEEE*, 2019, pp. 1–7. doi: 10.1109/ICCNT45670.2019.8944509.

- [22] A. Meneghetti, M. Sala, and D. Taufer, "A survey on pow-based consensus," *Annals of Emerging Technologies in Computing*, vol. 4, no. 1, pp. 8–18, Jan. 2020. doi: 10.33166/AETiC.2020.01.002.
- [23] H. Aissaua, M. Aliouat, A. Bounceur, and R. Euler, "A Distributed Consensus-Based Clock Synchronization Protocol for Wireless Sensor Networks," *Wirel. Pers. Commun.*, vol. 95, no. 4, pp. 4579–4600, Aug. 2017, doi: 10.1007/s11277-017-4108-4.
- [24] A. K. Yadav and K. Singh, "Comparative Analysis of Consensus Algorithms of Blockchain Technology," in *Advances in Intelligent Systems and Computing*, vol. 1097, 2020, pp. 205–218. doi: 10.1007/978-981-15-1518-7_17.
- [25] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, "A Survey of Distributed Consensus Protocols for Blockchain Networks," vol. 22, no. 2, pp. 1432–1465, 2020, doi: 10.1109/COMST.2020.2969706.
- [26] S. Velliangiri and P. Karthikeyan Karunya, *Blockchain technology: Challenges and security issues in consensus algorithm*. IEEE, 2020, pp. 1–8. doi: 10.1109/ICCCI48352.2020.9104132.
- [27] G. R. Carrara, L. M. Burle, D. S. V. Medeiros, C. V. N. de Albuquerque, and D. M. F. Mattos, "Consistency, availability, and partition tolerance in blockchain: a survey on the consensus mechanism over peer-to-peer networking," *Ann. des Telecommun. Telecommun.*, vol. 75, no. 3–4, pp. 163–174, Apr. 2020, doi: 10.1007/s12243-020-00751-w.
- [28] S. Alsaqqa and S. Almajali, "Blockchain Technology Consensus Algorithms and Applications: A Survey," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 15, p. 142, 2020, doi: 10.3991/ijim.v14i15.15893.
- [29] S. Pahlajani, A. Kshirsagar, and V. Pachghare, "Survey on Private Blockchain Consensus Algorithms," Apr. 2019, pp. 1–6. doi: 10.1109/ICHCT1.2019.8741353.
- [30] V. Gramoli, "From blockchain consensus back to Byzantine consensus," *Futur. Gener. Comput. Syst.*, vol. 107, pp. 760–769, Jun. 2020, doi: 10.1016/j.future.2017.09.023.
- [31] B. Lashkari and P. Musilek, "A Comprehensive Review of Blockchain Consensus Mechanisms," *IEEE Access*, vol. 9, pp. 43620–43652, 2021, doi: 10.1109/ACCESS.2021.3065880.
- [32] M. S. Ferdous, M. J. M. Chowdhury, and M. A. Hoque, "A survey of consensus algorithms in public blockchain systems for cryptocurrencies," *J. Netw. Comput. Appl.*, vol. 182, p. 103035, 2021, doi: <https://doi.org/10.1016/j.jnca.2021.103035>.
- [33] L. Ge, J. Wang, and G. Zhang, "Survey of Consensus Algorithms for Proof of Stake in Blockchain," *Secur. Commun. Networks*, vol. 2022, 2022.
- [34] H. Xiong, M. Chen, C. Wu, Y. Zhao, and W. Yi, "Research on Progress of Blockchain Consensus Algorithm: A Review on Recent Progress of Blockchain Consensus Algorithms," *Future Internet*, vol. 14, no. 2, 2022. doi: 10.3390/fi14020047.
- [35] A. Jain and D. S. Jat, "A Review on Consensus Protocol of Blockchain Technology BT - Intelligent Sustainable Systems," 2022, pp. 813–829.
- [36] S. Bano et al., *SoK: Consensus in the Age of Blockchains*. 2017.
- [37] K. Croman et al., "On Scaling Decentralized Blockchains," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9604 LNCS, Springer Verlag, 2016, pp. 106–125. doi: 10.1007/978-3-662-53357-4_8.
- [38] N. Chaudhry and M. M. Yousaf, "Consensus Algorithms in Blockchain: Comparative Analysis, Challenges and Opportunities," in *2018 12th International Conference on Open Source Systems and Technologies (ICOSSST)*, Dec. 2018, pp. 54–63. doi: 10.1109/ICOSSST.2018.8632190.
- [39] G. Bissias, B. N. Levine, A. P. Ozisik, and G. Andresen, *An Analysis of Attacks on Blockchain Consensus*. 2016.
- [40] S. Zhang and J.-H. Lee, "Double-Spending With a Sybil Attack in the Bitcoin Decentralized Network," *IEEE Trans. Ind. Informatics*, vol. 15, no. 10, pp. 5715–5722, Oct. 2019, doi: 10.1109/TII.2019.2921566.
- [41] D. Dasgupta, K. Datta Gupta, J. M. Shrein, • Kishor, and D. Gupta, "A survey of blockchain from security perspective," *J. Bank. Financ. Technol.*, vol. 3, no. 1, pp. 1–17, Apr. 2019, doi: 10.1007/s42786-018-00002-6.
- [42] M. Salimitari and M. Chatterjee, "A survey on consensus protocols in blockchain for IoT networks," *arXiv*. Sep. 2018.
- [43] C. Xu, K. Wang, and M. Guo, "Intelligent Resource Management in Blockchain-Based Cloud Datacenters," *IEEE Cloud Comput.*, vol. 4, no. 6, pp. 50–59, Nov. 2017, doi: 10.1109/MCC.2018.1081060.
- [44] S. Alam et al., "Blockchain-Based Solutions Supporting Reliable Healthcare for Fog Computing and Internet of Medical Things (IoMT) Integration," *Sustainability*, vol. 14, no. 22, p. 15312, 2022.
- [45] I. Bashir, *Mastering Blockchain: Deeper insights into decentralization, cryptography, Bitcoin, and popular Blockchain frameworks*. Packt Publishing Ltd, 2017.
- [46] F. Yang, W. Zhou, Q. Wu, R. Long, N. N. Xiong, and M. Zhou, "Delegated Proof of Stake With Downgrade: A Secure and Efficient Blockchain Consensus Algorithm With Downgrade Mechanism," *IEEE Access*, vol. 7, pp. 118541–118555, 2019, doi: 10.1109/access.2019.2935149.
- [47] D. Schuh, Fabian, Larimer, "Bitshares 2.0: general overview," p. 9, 2015, [Online]. Available: <https://cryptochainuni.com/wp-content/uploads/bitshares-general-overview.pdf>
- [48] R. Kotla, L. Alvisi, M. Dahlin, A. Clement, and E. Wong, "Zyzyva: Speculative Byzantine fault tolerance," *ACM Trans. Comput. Syst.*, vol. 27, no. 4, pp. 1–39, Dec. 2009, doi: 10.1145/1658357.1658358.
- [49] M. Castro and B. Liskov, "Practical Byzantine Fault Tolerance and Proactive Recovery," *ACM Trans. Comput. Syst.*, vol. 20, no. 4, pp. 398–461, Nov. 2002, doi: 10.1145/571637.571640.
- [50] A. N. Nikolakopoulos and J. D. Garofalakis, "NCDawareRank: A novel ranking method that exploits the decomposable structure of the web," in *WSDM 2013 - Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, 2013, pp. 143–152. doi: 10.1145/2433396.2433415.
- [51] S. Dziembowski, S. Faust, V. Kolmogorov, and K. Pietrzak, "Proofs of Space," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9216, Springer Verlag, 2015, pp. 585–605. doi: 10.1007/978-3-662-48000-7_29.
- [52] M. Ghosh, M. Richardson, B. Ford, and R. Jansen, *A TorPath to TorCoin: Proof-of-Bandwidth Altcoins for Compensating Relays*. 2014.
- [53] J. Sousa, A. Bessani, and M. Vukolic, "A byzantine Fault-Tolerant ordering service for the hyperledger fabric blockchain platform," in *Proceedings - 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2018*, Jun. 2018, pp. 51–58. doi: 10.1109/DSN.2018.00018.
- [54] Parity Technologies, "Proof of Authority - POA," 2017. <https://www.poa.network/for-users/whitepaper/poadao-v1/proof-of-authority>
- [55] I. Bentov, A. Gabizon, and A. Mizrahi, "Cryptocurrencies without proof of work," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9604 LNCS, pp. 142–157. doi: 10.1007/978-3-662-53357-4_10.
- [56] D. Puthal, S. P. Mohanty, V. P. Yanambaka, and E. Kougianos, "PoAh: A Novel Consensus Algorithm for Fast Scalable Private Blockchain for Large-scale IoT Frameworks," *arXiv*, pp. 1–26, Jan. 2020.
- [57] S. P. Mohanty, V. P. Yanambaka, E. Kougianos, and D. Puthal, "PUFchain: A Hardware-Assisted Blockchain for Sustainable Simultaneous Device and Data Security in the Internet of Everything (IoE)," *IEEE Consum. Electron. Mag.*, vol. 9, no. 2, pp. 8–16, Mar. 2020, doi: 10.1109/MCE.2019.2953758.
- [58] D. H. Kim, R. Ullah, and B. S. Kim, "RSP Consensus Algorithm for Blockchain," 2019 20th Asia-Pacific Netw. Oper. Manag. Symp. Manag. a Cyber-Physical World, APNOMS 2019, pp. 1–4, 2019, doi: 10.23919/APNOMS.2019.8893063.
- [59] J. Yang, M. M. H. Onik, N. Y. Lee, M. Ahmed, and C. S. Kim, "Proof-of-familiarity: A privacy-preserved blockchain scheme for collaborative medical decision-making," *Appl. Sci.*, vol. 9, no. 7, p. 1370, Apr. 2019, doi: 10.3390/app9071370.
- [60] J. Zou, B. Ye, L. Qu, Y. Wang, M. A. Orgun, and L. Li, "A Proof-of-Trust Consensus Protocol for Enhancing Accountability in Crowdsourcing Services," *IEEE Trans. Serv. Comput.*, vol. 12, no. 3, pp. 429–445, 2019, doi: 10.1109/TSC.2018.2823705.

- [61] M. Milutinovic, W. He, H. Wu, and M. Kanwal, "Proof of Luck," in Proceedings of the 1st Workshop on System Software for Trusted Execution, Dec. 2016, pp. 1–6. doi: 10.1145/3007788.3007790.
- [62] N. Lasla, L. Al-Sahan, M. Abdallah, and M. Younis, "Green-PoW: An energy-efficient blockchain Proof-of-Work consensus algorithm," *Comput. Networks*, vol. 214, p. 109118, Sep. 2022, doi: 10.1016/J.COMNET.2022.109118.
- [63] M. Du et al., "A review on consensus algorithm of blockchain," in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2017, vol. 2017-Janua, pp. 2567–2572. doi: 10.1109/SMC.2017.8123011.
- [64] X. Li, P. Jiang, T. Chen, X. Luo, and Q. Wen, "A survey on the security of blockchain systems," *Futur. Gener. Comput. Syst.*, vol. 107, pp. 841–853, Jun. 2020, doi: 10.1016/j.future.2017.08.020.
- [65] S. K. Kim and J. H. Huh, "A study on the improvement of smart grid security performance and blockchain smart grid perspective," *Energies*, vol. 11, no. 8, p. 1973, Jul. 2018, doi: 10.3390/en11081973.
- [66] D. Larimer, "Transactions as proof-of-stake," *Cryptochainuni.Com*, pp. 1–8, 2013, [Online]. Available: <https://cryptochainuni.com/wp-content/uploads/Invictus-Innovations-Transactions-As-Proof-Of-Stake.pdf>
- [67] I. Bentov, C. Lee, A. Mizrahi, and M. Rosenfeld, "Proof of Activity: Extending Bitcoin's Proof of Work via Proof of Stake," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, no. 3, pp. 34–37, Dec. 2014, doi: 10.1145/2695533.2695545.
- [68] D. Mazières, "The Stellar Consensus Protocol A federated model for Internet-level consensus," *pdfs.semanticscholar.org*, 2015.

A Novel Approach for Identification of Figurative Language Types in Devanagari Scripted Languages

Jatinderkumar R. Saini, Preety Sagar, Hema Gaikwad

Symbiosis Institute of Computer Studies and Research, Symbiosis International (Deemed University) Pune, India

Abstract—Poetry can be defined as a form of literary expression that uses language and artistic techniques to evoke emotions, create imagery, and convey complex ideas in a concentrated and imaginative manner. It is a form of written or spoken art that often incorporates rhythm, meter, rhyme, and figurative language to engage the reader or listener on multiple levels. There is no automated system that can identify figures of speech (FoS) in poetry using Natural Language Processing (NLP) methods. In this research paper, the authors categorized four types of FoS: व्रत्या अनुप्रास (Type of alliteration), छेकानुप्रास (Type of alliteration), अन्तल्यानुप्रास (Rhyme), and पुनरुक्ति (Repetition) using two custom algorithms, Koshur and Awadhi (KA and AA), developed specifically for three different language corpora of poems (Koshur (K), Awadhi (A), and Hindi (H)). To evaluate the effectiveness of these algorithms, the authors conducted tests on three languages using four distinct approaches: with stopwords without optimization, with stopwords with optimization, without stopwords without optimization, and without stopwords with optimization. Authors have put lots of effort into identifying FoS in not only one single language but in three Devanagari scripted languages. This research work is the first of its kind. The average accuracy without stopwords was not up to the mark. The authors then optimized both algorithms and again tested them on the same corpora with or without stopwords, resulting in a significant increase in accuracy.

Keywords—Figures of speech (FoS); natural language processing (NLP); Koshur; Awadhi

I. INTRODUCTION

India is renowned for its rich cultural heritage and vast literary traditions. With twenty-two scheduled languages, the country boasts a remarkable diversity of literature, and poetry stands out as one of its prominent forms of expression. Indian poetry, rooted in ancient times, has flourished through the ages, reflecting the myriad influences of its diverse regions and communities.

Poetry is characterized by its unique use of language, employing words and phrases chosen for their sound, texture, and connotation, as well as their literal meaning. Poets often strive to capture the essence of an experience or emotion, using vivid imagery and metaphorical language to paint a picture or evoke sensory impressions. While poetry can take many forms and styles, ranging from traditional structures like sonnets and haikus to free verse and experimental formats, it is united by its emphasis on creative expression, aesthetic beauty, and the power of words. It explores themes such as love, nature, human experiences, social and political issues, and the mysteries of existence. To acquire a deeper knowledge of a given poetry, we

might focus on numerous elements. These aspects include Emotion (Rasa), voice, diction, imagery, figures of speech, syntax, sound, rhythm, and verse meter [1]. These unfold the poem's actual theme, meaning, or feelings. The figure of speech (FoS) is one such element that creates a mesmerizing effect and produces magic. They are considered the ornaments used for the beautification of the poems.

This is a maiden attempt by the authors of this research paper to identify four different forms of figures of speech (FoS), namely व्रत्या अनुप्रास 'Type of alliteration', छेकानुप्रास 'Type of alliteration', अन्तल्यानुप्रास 'Rhyme' and पुनरुक्ति 'Repetition' (which is an implicit type of alliteration) on three different language corpora Koshur, Awadhi, and Hindi, implemented in two different algorithms Koshur and Awadhi algorithm (KA and AA) and later optimized with and without stopwords. All three languages, Koshur Awadhi, and Hindi are based on the Devanagari script.

Kashmiri or commonly referred to as 'Koshur', is a branch of the Indo-Aryan language in the linguistic landscape of India spoken by about seven million people from the state of Jammu and Kashmir, India. The Perso-Arabic script, the Sharda script, and the Devanagari script are the three main scripts used to write Koshur [2,3]. The Perso-Arabic script is considered the official script of Koshur, and Devanagari is used by Kashmiri Hindus for literature purposes [4]. The pioneers of literature, the Kashmiri poets, brought freshness and sweetness to the poems using this script [5]. The corpora considered in this research paper were Koshur poems written in Koshur Devanagari script.

Awadhi is spoken in twenty districts of India and eight districts of Nepal [6]. Awadhi is a language spoken in the Awadh region of Uttar Pradesh and some parts of north India. Awadhi is one of the dialects of Hindi, and it has 38 million native speakers [7]. But very less research work has been performed on the Awadhi language. Prominent texts like Ramcharitmanas, one of the most important religious texts in the world, is written in Awadhi. The Epic Ramcharitmanas is traditionally divided into several major kaandas or books, that deal chronologically with the major events in the life of Rama—Bala Kaanda, Ayodhya Kaanda, Aranya Kaanda, Kishkindhaa Kaanda, Sundara Kaanda, Lanka Kaanda, and Uttara Kaanda. Ramcharitmanas has seven Kaanda such as the other text like Hanuman Chalisa and Padmavat are also written in Awadhi. [8][9][10]. Today, the world is rapidly heading towards digitization. e-Books are easy to publish, handle and promote. All the important ancient books and texts are being digitized.

Hindi is the official language of India and holds a significant place in the country's cultural and literary heritage. Hindi belongs to the Indo-Aryan language family and shares roots with other languages, such as Sanskrit. Poetry in Hindi is a rich and diverse art form that has flourished for centuries. It encompasses various forms and styles, including Bhakti poetry, Ghazals, Doha, and Geet. Each form has its unique structure, rhythm, and thematic focus.

Automated identification of four types of FsoS in Devanagari scripted three major languages, Koshur, Awadhi, and Hindi, are first of its kind, and hybridization of two separate algorithms further improved the performance. No such work, as per authors' best-known knowledge, is seen with special reference to three languages. The author's contribution to this research covers the following items: -

- Identification of four different FsoS, using two different algorithms and their implementation on three different language corpora.
- Optimization of algorithms with and without stopwords.
- Comparison of both KA and AA algorithms considering accuracy and execution time taken by each algorithm.

The remaining sections of the paper are organized as follows: Section II gives a detailed literature review of the work. Section III explains the methodology used along with the algorithms. Section IV shows the interpretation of the results. Section V concludes the work and discusses its future scope.

II. LITERATURE REVIEW

A thorough literature review was done to gather information about the present state of the research effort in the field of Natural Language Processing (NLP) which is a branch of Artificial Intelligence and Machine Learning, and work done in FoS in poetry. Research in poetry and nearby segments like poetry generation, sentiment analysis, text classification, meter classification, etc., has been seen in many Indian and foreign languages like Marathi, Punjabi, Hindi, Arabic, Chinese, Persian, etc., but still, many Indian regional languages are either completely absent or badly represented on the NLP map. Although authors can find some work in the Hindi language, but Koshur and Awadhi language are ones in which very little or no study has been done.

According to Chopra et al. [11] NLP is a field of study in computer science, artificial intelligence, and linguistics that investigates how computers interact with human or natural languages. NLP is primarily concerned with human-computer interaction. NLP was also felt necessary because computers could access much information recorded or stored in natural language. Saini and Kaur [1] worked on poetry characteristics such as diction, rhyme, and rhythm, setting it apart from other genres of literature. These factors were experimented with, to attempt an automatic system for categorizing poetries based on emotional states tested to develop a system for categorizing poetries based on the Indian concept of 'Navrasa.' Kushwah and Joshi [12] investigated Hindi poetry based on 'Chhand',

one of the properties of Hindi poems. They created an algorithm that detects the presence of 'Rola Chhand' in any poem provided as input. A few poems are available in digital form, but their poetic properties are not aimed at, and their algorithm focuses on one such property.

Audichya and Saini [13] worked on producing automatic metadata for 'Chhand' based on the stanzas of the poems. They also provided superior techniques for metadata creation and procedures for 'Muktak Chhands'. It was the first time that not only rules of the 'Chhands' were identified but also were confirmed and modelled from the standpoint of computational linguistics. Audichya and Saini [14] worked on identifying three Hindi figures of speech using NLP. They also created a systematic structure of types and sub-types of Hindi FoS. Bafna and Saini

The study [15] recovered tokens from two corpora using two different methodologies. To count and contrast extracted tokens, BaSa, and Zipf's law were employed. Further token comparison between the two approaches is accomplished. They used both Hindi and Marathi poems and prose. To demonstrate that Hindi and Marathi behave similarly for NLP operations, common tokens from corpora of Marathi and Hindi poetry and prose were identified. It was established that BaSa outperforms Zipf's law. Kaur and Saini [16] worked on creating a content-based classifier for Punjabi poetry. After going through the pre-processing layer, more than 2,034 poems and 31,938 tokens were separated and weighted using the term frequency (TF) and term frequency-inverse document frequency (TF-IDF).

Pal and Patel [17] classified poetry based on nine different types of Rasas like Shringar, Hasya, Rudra etc., and used a mix of part-of-speech and emotion-based features to classify poems into different types. In research by Lone et al. [18], a Kashmiri-to-English Machine Translation System was presented, as well as it highlighted various features of the Kashmiri language. Their method was built on machine intelligence, and it can learn various translation rules from a series of translated input words by employing Long Short-term Memory (LSTM) architecture for deep sequence learning. The paper also reports difficulties and challenges associated with the work. Mir and Laway [19] worked on Word Sense Disambiguation (WSD) System for Kashmir Language; they designed Sense Annotated Corpus for Kashmiri Language and WSD Data Set. Ahmad and Syam [20] developed a Parts-of-Speech tagger (POS) in Perso-Arabic script for the Kashmiri language with an accuracy of 80.64%. Rasool et al. [21] opened the doors to the creation of a powerful multilingual machine translation system that includes Kashmiri as one of the languages. The aim of the researcher here is to incorporate Kashmiri into UNL (Universal Networking Language) framework. In this work, a selected Kashmiri corpus is analysed to UNL using IAN, and subsequently, Kashmiri expressions are generated from UNL expressions using EUGENE. Gilkar et al. [22] proposed a POS rule-based tagger for Kashmiri written in the Nastaliq script. The authors tried to create an automatic tagger for Kashmiri corpora using a rule-based and stochastic (hybrid) tagging approach. Aabid et al. [23] presented an Automatic Recognition System (ARS) that allows computers to understand natural

speech for recognizing Kashmiri digits zero (sefar) to nine (nov) spoken in isolation by different male and female speakers. In study by Ramakrishna et al. [24] Kashmiri palatalized consonants were examined, which were related to, the i-matra vowel and palatalized in Kashmiri phonology. For the purpose of analysing Kashmiri voice data.

Firdaus et al. [25] explained that poem writing is a way to express ideas, thoughts, and feelings using artful language. Writing poetry is a skill that is taught in schools. There are hardly any students who are interested in poetry. Less appealing is the poetry writing process used in schools. As a result, the purpose of this study is to demonstrate an original way to write poetry using the "Atafora" technique by fusing students' sensory experiences with a metaphorical figure of speech. Malik et al. [26] discussed the figures of speech in the songs from Rose's album "R". The study's objective was to characterize four different types of figures of speech used in the two-track lists of the R album: (a) comparative figures of speech, (b) contradiction figures of speech, (c) affirmation figures of speech, and (d) satirical figures of speech. This study used a qualitative research design and an analytical framework. The -R- album's two songs with the title "On the Ground and Gone" that use figure of speech are the study's source of data. Krishna et al. [27] stated that poets frequently use figurative language to convey their thoughts and emotions. Poetry that captures the attention of the readers will be enhanced by figurative language. Paradox is frequently used in figurative poetry to emphasize the poem's message. This study focuses on the implications of paradox in the poem by Rudyard Kipling. The authors used the descriptive qualitative method to examine Rudyard Kipling's chosen poems' figure of speech. The study discovered 17 paradoxes, including seven rhetorical paradoxes, seven social paradoxes, two logical paradoxes, and one philosophy of science contradiction. Maula et al. [28] identified the figurative language in Lady Gaga's album titled "Always Remember Us This Way" using descriptive qualitative research. Wati et al. [29] discussed the comparative figurative language in the poems from Emi Suy's poetry collection titled Ibu Menanak Nasi hingga Matang Usia Kami. The anthology was published in 2022, served as the study's data source. Heuristic reading, a first-level semiotic reading technique, and Hermeneutic reading, a second-level structuralism-semiotic technique, was used for data collection and data analysis purpose, respectively. Naaz et al. [30] discussed different tools such as Text2Matr, RPaGen, and FoSCal. The Text2Matr tool offers a chanda-related observation, including chanda type detection and classification, rhythm determination, chanda correctness verification, etc. For an input Hindi poetry, the RPaGen tool provides rhyme pattern(s). The tool FoSCal, used for generating alankara scores, and the tool's FoSCal uses the pattern created by RPaGen for alankara rating. Mahdi et al. [31] explained that poets use a wide range of writing strategies when creating new poems. This paper discussed the use of personification, symbolism, and figure of speech such as Simile and Metaphor in British love poems and lyrics. Nidi et al. [32] discussed the personification figure of speech and its general meaning. The two varieties of personification FsoS prosopography and prosopopeia are used. The author has used the poems by Robert Frost.

Setiani et al. [33] stated that poetry is a literary form that includes stanzas, lines, rhythms, and rhymes. Denotative and connotative meanings are frequently utilized in poetry. Denotative means emotional feeling of the word, and connotative means figurative language. Hutaurok et al. [34] mentioned the importance of figures of speech. The personification and apostrophe create pictures in the mind of reader or listener, and pictures help to convey the message faster than words. Sayakhan et al. [35] Figurative language is used frequently in ordinary conversation, popular music, television, and commercial topics as well as in classic works like Shakespeare and the Bible. The author has specially discussed two figures of speech such as personification and apostrophe. A person is just mentioned in an apostrophe, whereas with personification, inanimate objects are given human characteristics.

While performing the literature review the authors found few papers on FoS but no papers were found in Koshur and Awadhi.

Therefore, the gap with the authors was huge, the reason behind this gap is mainly that the languages are highly resourceless, it is tedious to work with such resourceless languages and researchers have to face lots of difficulties as no initial work has been carried out before.

To address this significant gap, the authors have embarked on developing FoS algorithms. The research question then becomes: How can the development of FoS algorithms alleviate the challenges posed by the lack of resources in these languages and contribute to narrowing this gap?

For the above mentioned research question authors have worked on the objectives like development, optimization and implementation of FoS algorithms for Devanagari scripted languages.

III. METHODOLOGY

To identify FsoS in three different languages, authors developed the logic represented in Fig. 1. The input to the flowchart was the poems in all three languages. The entered poems then need to be tokenized on the basis of sentences, words, and letters, and accordingly, rules were applied as per the type of FoS identified.

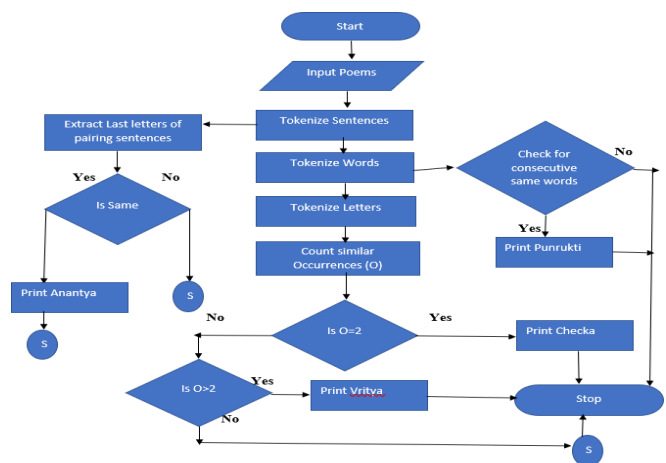


Fig. 1. Logical representation.

TABLE I. SUMMARY OF FoS, TRANSLITERATION, TRANSLATION, DEFINITION, AND EXAMPLE

FoS- N	T	RST	D	H	E	K	A
त्रया अनुप्रास	Vrity Anupras (VA)	Type of Alliteration	consonant word repeats more than once.	चारुचंद्र की चंचल किरणें, खेल रहीं हैं जल थल में	Serve Fast food with fast speed	यि छु म्योन वतन यि छु चोन वतन	बंदउ गुरु पद पदुम परागा सुरुचि सुबास सरस अनुरागा
छेकानुप्रास	Cheka Anupras (CA)	Type of Alliteration	The same letter used repeatedly but only 2 times	फूले फले राष्ट्र दिन-रात मेरा	My country, my nation is blooming	वदुन रिवुन पँतिम ति बार याद छुम में याद छुम	जो सुमिरत सिधि होइ गन नायक करिबर बदन
अन्त्यानुप्रास	Anantya Anupras (AA)	Rhyme	Rhyming pattern at the end of lines	बुंदेले हरबोलों के मुँह हमने सुनी कहानी थी।खुब लडी मर्दानी वह तो झाँसी वाली रानी थी	On the fifteenth of May, in the jungle of nool, In the heat of the day, in the cool of the pool	सुमरन पनन्य दिचोनम, लोलुक निशान वेसिये।रँछरुन तोगुम न रोवुम, ओसुम न बान वेसिये	एहि महँ रघुपति नाम उदारा। अति पावन पुरान श्रुति सारा
पुनरुक्ति अलंकार	Punrukti (PU)	Repetition	one word used twice consecutively without changing its meaning	मीठा-मीठा रस टपकता	milky-milky chocolate, covered with coffee beans.	जिंदगी हुँद मोदुर-मोदुर फरागथ.	बार बार रघुबीर सँभारी। तरकेउ पवनतनय बल भारी॥

Legends- FoS N-FoS Name,, T-, Transliteration , RST-Roman Scripted Translation, D-Definition, H Hindi, ,E-English-Koshur ,A-Awadhi

TABLE II. METADATA OF THREE DIFFERENT LANGUAGE POEMS

S.No.	Attributes	Koshur	Awadhi	Hindi
1	CT	DA	RAM	ABV
2	NoP	17	7	17
3	CS	341	340	353
4	ExT	2096	2819	1716
5	ExSW	510	264	862

Legends-Corpus Type- CT, DA-Different Authors, RAM-Ramcharitmanas, ABV-Poems of Shri Atal Bihari Vajpayee, NoP-Number of Poems, CS-CorpusSize, ExT-Extraction of tokens, ExSW-Extraction of Stop Words.

To detect or identify the figures of speech in Koshur, Hindi, and Awadhi, authors need detailed descriptions of all four types of FoS and their rules. Table I shows the definitions and examples of FoS in all four languages.

The Table II above shows the details of corpus used. Well-known poems from three different languages were taken. Awadhi poems were extracted from different kaand of epic Ramcharitmanas where one kand represents one section. Throughout the paper, the authors used the word section in place of kand. For Hindi, poems of Shri. Atal Bihari Vajpayee ji had been taken, and poems of different authors were extracted for Koshur. The selection of Koshur poetry was made in order to include a variety of poems, including those about love, nature, patriotism, and other topics. For the implementation of the algorithm, lines of the poems were considered as Units of Measurement (UoM). In Awadhi, 40-50 lines were taken from each section. For Hindi and Koshur, seventeen poems were considered having 353 and 341 lines, respectively. Tokenization is a process of breaking sentences into words. After applying the tokenization process, the extracted words were 2819, 1716, and 2096 for Awadhi, Hindi, and Koshur, respectively. In NLP, stopwords are words that have less relevance and do not carry valuable information. The authors extracted 264, 862, and 510 stopwords in Awadhi, Hindi, and Koshur. Awadhi and Koshur are low-resource languages, and there are various challenges, such as the non-availability of linguistic documents like dictionaries, wordnet, thesaurus,

stopwords list, POST etc. So, the extraction of stopwords was also a challenge for the authors.

As all three languages support the Devanagari script, the Hindi stopwords list has been taken as the benchmark. To identify Koshur stopwords, the authors have considered three lists of stopwords from three different languages (Hindi, English, and Punjabi). Some new words were also added to the list that were fetched from credible sources. For the identification of Awadhi stopwords, the authors have referred list of Hindi stopwords and modified them according to the Awadhi language. The authors also contributed new words from the literature of Awadhi.

IV. RESULT AND DISCUSSION

The authors developed two different algorithms, one for the Koshur language and another for Awadhi. Both algorithms were then implemented and tested on the corpus with stopwords and without stopwords. Optimization of algorithms was also done to improve the performance. Algorithms developed for the Koshur language were implemented in Hindi and Awadhi, and the algorithms developed for the Awadhi language were implemented in both Koshur and Hindi.

The performance of the Koshur and Awadhi algorithms on a corpus with stopwords and without optimization is shown in Table III. Table IV shows the performance of both algorithms with stopwords and with optimization.

TABLE III. IMPLEMENTATION OF KOSHUR AND AWADHI FoS ALGORITHMS IN THREE LANGUAGES WITH STOPWORDS WITHOUT OPTIMIZATION

FoS	KA						AA					
	K	ET	A	ET	H	ET	K	ET	A	ET	H	ET
VA	85.71	0.09	83.33	0.12	80.00	0.03	100.00	0.02	97.60	0.03	80.00	0.01
CA	93.33	0.08	86.66	0.25	94.11	0.06	96.40	0.01	97.10	0.04	94.80	0.02
AA	92.30	0.15	94.11	0.60	94.11	0.06	73.90	0.01	90.30	0.02	74.10	0.01
PU	91.66	0.15	92.30	0.60	92.30	0.06	96.10	0.03	100.00	0.01	100.00	0.01
Avg	90.75	0.12	89.10	0.39	90.13	0.05	91.60	0.02	96.25	0.03	87.23	0.01

Legend-ET-Execution Time

TABLE IV. IMPLEMENTATION OF KOSHUR AND AWADHI FoS ALGORITHMS IN THREE LANGUAGES WITH STOPWORDS AND WITH OPTIMIZATION

FoS	KA						AA					
	K	ET	A	ET	H	ET	K	ET	A	ET	H	ET
VA	100.00	0.66	90.00	0.65	100.00	0.12	100.00	0.02	95.20	0.02	90.00	0.01
CA	93.33	0.70	100.00	0.83	94.11	0.15	98.30	0.05	98.80	0.04	96.80	0.03
AA	92.30	0.71	92.85	1.67	94.11	0.40	74.80	0.01	98.63	0.02	75.30	0.02
PU	91.66	0.70	80.00	1.67	93.30	0.40	96.20	0.02	100.00	0.01	100.00	0.01
Avg	94.32	0.70	90.71	1.21	95.38	0.27	92.33	0.03	98.16	0.02	90.53	0.02

TABLE V. IMPLEMENTATION OF KOSHUR AND AWADHI FoS ALGORITHMS IN THREE LANGUAGES WITHOUT STOPWORDS WITHOUT OPTIMIZATION

FoS	KA						AA					
	K	ET	A	ET	H	ET	K	ET	A	ET	H	ET
VA	83.33	0.12	83.00	0.27	80.00	0.15	87.50	0.02	100.00	0.02	100.00	0.01
CA	53.33	0.12	86.66	0.36	58.82	0.15	98.76	0.01	97.25	0.02	100.00	0.01
AA	92.30	0.32	92.85	0.75	92.30	0.32	100.00	0.01	98.70	0.02	100.00	0.01
PU	91.66	0.31	80.00	0.74	92.30	0.31	85.71	0.01	77.78	0.01	83.33	0.01
Avg	80.16	0.22	85.63	0.53	80.86	0.23	92.99	0.01	93.43	0.02	95.83	0.01

TABLE VI. IMPLEMENTATION OF KOSHUR AND AWADHI FoS ALGORITHMS IN THREE LANGUAGES WITHOUT STOPWORDS WITH OPTIMIZATION

FoS	KA						AA					
	K	ET	A	ET	H	ET	K	ET	A	ET	H	ET
VA	100.00	0.14	83.33	0.06	100.00	0.06	87.50	0.02	100.00	0.01	100.00	0.01
CA	93.33	0.15	94.11	0.13	94.11	0.11	97.93	0.01	99.09	0.01	100.00	0.01
AA	92.30	0.31	94.11	0.12	94.11	0.11	100.00	0.01	99.34	0.02	100.00	0.01
PU	91.66	0.30	92.30	0.12	92.30	0.12	85.71	0.01	77.78	0.01	83.33	0.01
Avg	94.32	0.23	90.96	0.11	95.13	0.10	92.79	0.01	94.05	0.01	95.83	0.01

The performance of the Koshur and Awadhi algorithms on a corpus without stopwords and without optimization is shown in Table V and Table VI shows performance without stopwords with optimization.

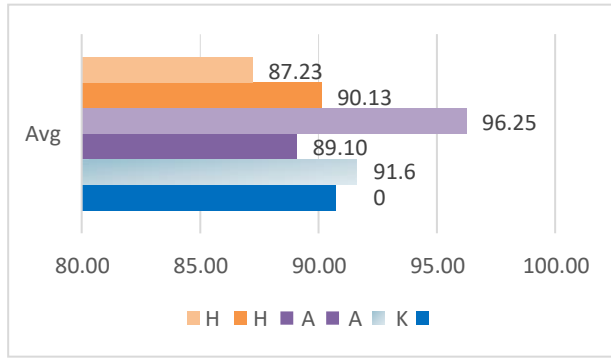
Fig. 2 shows the average accuracy of two algorithms with stopwords, with and without optimization. Fig. 4 show the result of the Koshur and Awadhi algorithms without stopwords, with and without optimization. The dark shades represent the result of Koshur algorithm and light shades represent the result of Awadhi algorithm.

The authors' observation on Fig. 2 is that both algorithms have increased the performance after optimizing the existing

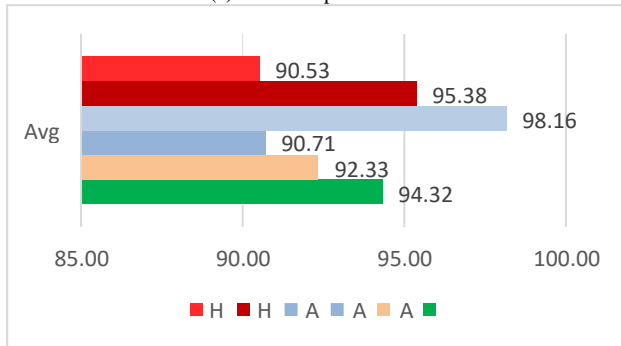
code and removal of the extra loops and unwanted instructions from the code. Fig. 3 displays the time taken by the algorithms.

There are four types of approaches considered by the authors the names are WSWWO-With Stopwords With Optimization, WSWWO-Without Stopwords Without Optimization, WOSWOO-Without Stopwords With Optimization, WOSWOO-Without Stopwords Without Optimization. The graph shows the relation between ET and the approaches. The time taken by the algorithms before optimization was little high than the time taken after optimization. The orange colour line shows with stopwords with optimization, and blue colour represents with stopwords

without optimization.



(a) Without optimization.



(b) With optimization.

Fig. 2. (a,b) Koshur and Awadhi algorithm accuracy with stopwords.



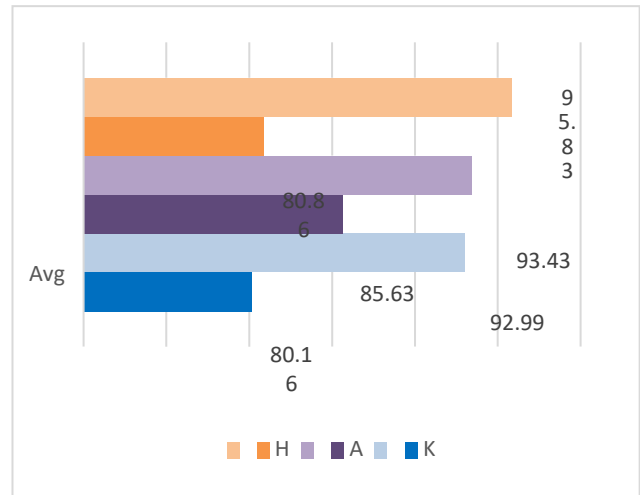
Fig. 3. ET taken by algorithms with stopwords.

The observation related to Fig. 4 is that after removing the stopwords, the accuracy reduces noticeably. The stopwords are words that do not make semantic sense and are unwanted for processing. As we know that poets use figurative language for writing poems and create a magical effect, but when we remove the stopwords, this effect breaks the rhythm of the poem, and hence code is not able to catch the particular FoS. So our contribution is improvised algorithms that increased the accuracy of the previous algorithms without stopwords up to a great extent. Fig. 5 displays the execution time taken by both the algorithms.

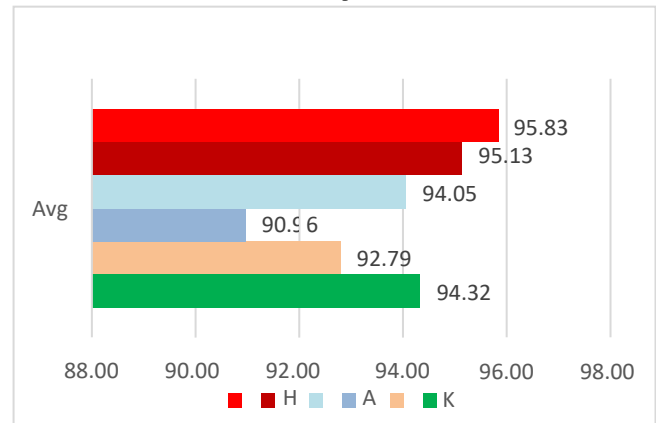
The algorithms after optimization showed the difference in timing. The optimized algorithms were taking less time than the algorithms without optimization.

Fig. 6 show the accuracy of individual FoS named VA, CA, AA, and PU. The blue, pink, and green colours represent

Koshur, Awadhi, and Hindi languages. The Koshur algorithm with stopwords was giving 100% accuracy in Hindi and Koshur languages for VA FoS. The same algorithm was giving 100% accuracy in Awadhi in CA FoS. The accuracy for AA and PU FoS in Hindi was 94.11% and 93.30, which was the highest among all three languages. The same algorithm, when used on corpus without stopwords, was giving 100% results in Koshur and Hindi for VA FoS. The CA and AA FoS scored 94.11 % accuracy in Awadhi and Hindi. The 92.30% accuracy was received in Awadhi and Hindi languages for PU FoS.



(a) Without optimization.

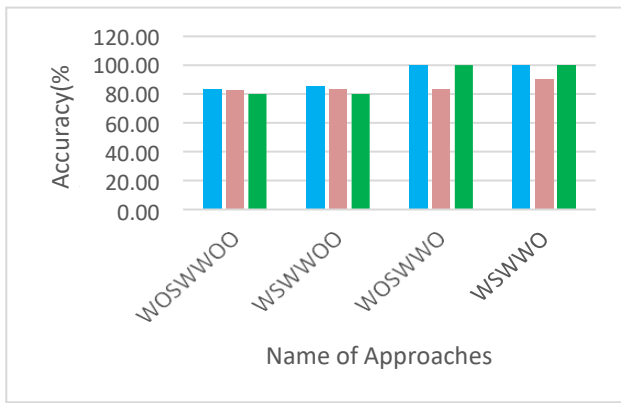


(b) With optimization.

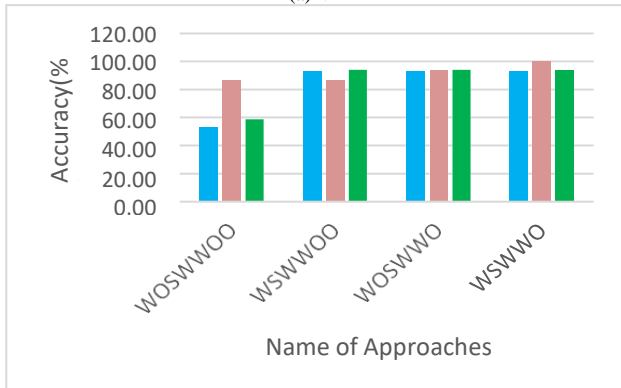
Fig. 4. (a,b) Koshur and Awadhi algorithm accuracy without stopwords.



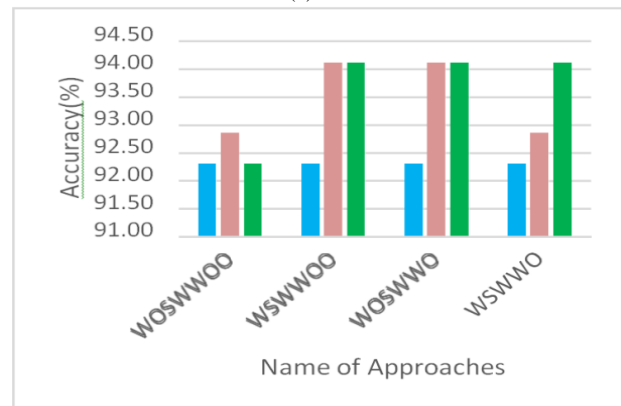
Fig. 5. ET taken by algorithms without stopwords.



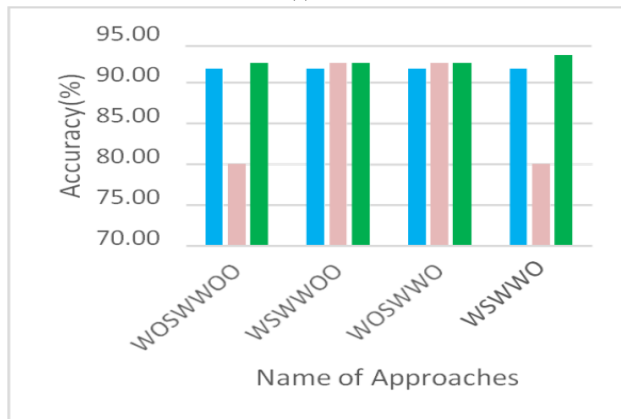
(a) VA



(b) CA

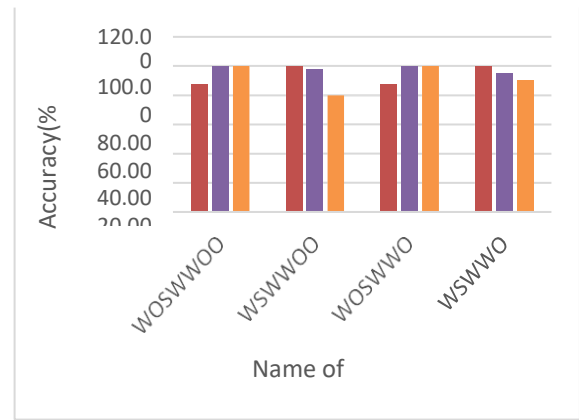


(c) AA

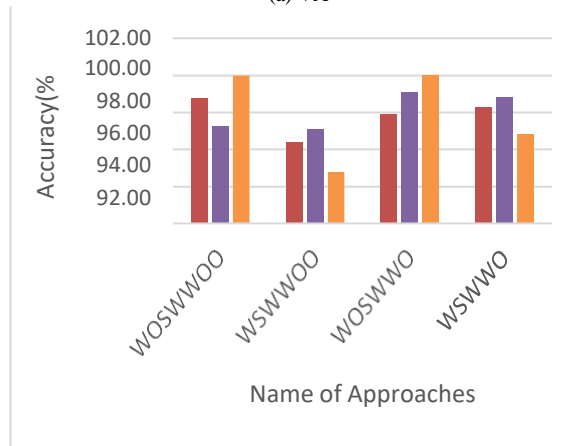


(d) PU

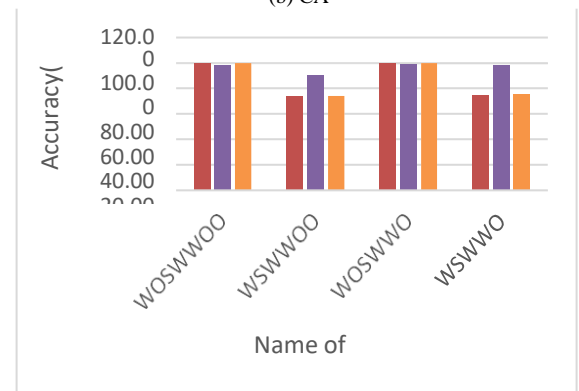
Fig. 6. Application of Koshur algorithm for four FsoS.



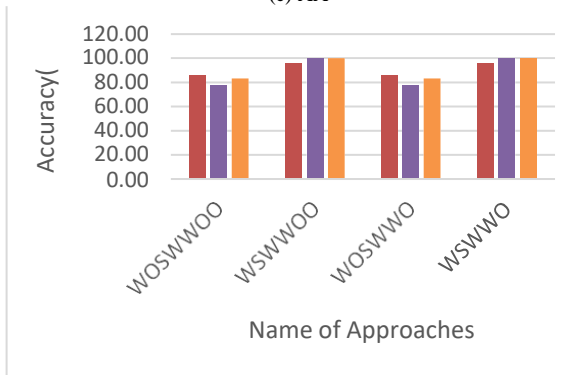
(a) VA



(b) CA



(c) AA



(d) PU

Fig. 7. Application of Awadhi algorithm for four FsoS.

Similarly, Fig. 7 show the individual accuracy of FoS named VA, CA, AA, and PU. The dark red, violet and orange colours are used for Koshur, Awadhi, and Hindi. While implementing the Awadhi algorithm on corpora with stopwords, the authors recorded 100% accuracy in VA and PU FoS in Koshur, Hindi, and Awadhi languages, respectively. For CA and AA, the accuracy was 98.80 % and 98.63% in Awadhi. The Awadhi algorithm without stopwords provided 100% accuracy in VA in Awadhi and Hindi. The same algorithm gives 100 % results in Hindi for CA and AA FoS. For the Koshur language, 85.71% accuracy was scored in PU FoS.

After analyzing the results the authors derived that the initial implementation of both algorithms performed less accurately on datasets without stopwords compared to datasets that included stopwords. The disparity in accuracy was quite noticeable. However, after applying optimization techniques to both the KA and AA algorithms, there was a significant enhancement in accuracy for both types of datasets – with and without stopwords. The optimization process played a crucial role in enabling the authors to achieve higher accuracy levels for their algorithms, even when stopwords were removed from the text. This underscores the effectiveness of the optimization in improving the algorithms' performance across different types of textual data.

V. CONCLUSION AND FUTURE SCOPE

The Koshur, Awadhi, and Hindi corpora were used by the authors of this research study to identify four different forms of figures of speech. Two algorithms, KA and AA, were applied by the authors on a corpus with and without stopwords. Authors then improved the algorithms' performance and applied them again on a corpus with and without stopwords.

Both the algorithms before optimization, when implemented on corpora without stopwords, showed low accuracy compared to corpora with stopwords. The difference in the accuracy was quite noticeable. After implementing optimized algorithms (both KA and AA), a significant increase in accuracy can be seen on the corpus without stopwords and with stopwords as well. Optimization helped authors to increase the accuracy of algorithms even after removing the stopwords.

This accomplishment offers a tangible solution to the stark resource limitations these languages confront. The optimized algorithms exhibit a capacity to extract meaningful insights from text, compensating for the lack of extensive linguistic resources. Consequently, the development of these FoS algorithms not only provides a means to navigate the complexities of resource-scarce languages but also contributes significantly to narrowing the existing accuracy gap between corpora with and without stopwords.

In the future, the work will be extended for other Devanagari scripted languages, and the identification of more FoS can be explored and tested.

REFERENCES

- [1] Saini, J. R., & Kaur, J. (2020). Kāvi: An Annotated Corpus of Punjabi Poetry with Emotion Detection Based on 'Navrasa'. *Procedia Computer Science*, 167,1220-1229. <https://doi.org/10.1016/j.procs.2020.03.436>
- [2] <https://prindia.org/billtrack>, "The-jammu-and-kashmir-official-languages-bill-2020", available online, accessed on (07-07-23).
- [3] <https://www.indiacode.nic.in/>, "The Jammu and Kashmir official Language Act, 2020" available online, accessed on (07-07-23).
- [4] Why we demand Devanagari as additional/co- script for Kashmiri language, by Dr. R.K. Bhat <http://ikashmir.net/rkbhat/script.html>
- [5] Handoo, J. (1979). *Contemporary Kashmiri Poetry*. *Indian Literature*, 22(5), 145-154.
- [6] <https://www.omniglot.com/writing/awadhi.html>, "online encyclopedia of writing systems & Languages", available online, accessed on (10-7-23)
- [7] https://archive.org/details/in.ernet.dli.2015.3210_6/page/n23/mode/2up
- [8] <http://www.ramcharitmanas.org/>, "About Ramcharitmanas", available online, accessed on (02-7-23)
- [9] <http://aumamen.com/stotra/hanuman-chalisa>, "Aumamen", available online, accessed on (11-07-23)
- [10] https://kavitakosh.org, "Kavita Kosh", available online, accessed on (11-07-23).
- [11] Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. *International journal of technology enhancements and emerging engineering research*, 1(4), 131-134.
- [12] Kushwah, K. K., & Joshi, B. K. (2017). Rola: AnEqui-Matrik Chhand of Hindi Poems. *Int J Comput Sci Inf Secur (IJCSIS)*, 15(3).
- [13] Audichya, M. K., & Saini, J. R. (2021). Stanza type identification using systematization of versification system of Hindi poetry. *International Journal of Advanced Computer Science and Applications*, 12(1).
- [14] Audichya, M. K., & Saini, J. R. (2021). Towards natural language processing with figures of speech in Hindi poetry. *International Journal of Advanced Computer Science And Applications*, 12(3).
- [15] Bafna, P. B., & Saini, J. R. (2020). AnApplication of Zipf's Law for Prose and Verse Corpora Neutrality for Hindi and Marathi Languages. *International Journal of Advanced Computer Science and Applications*, 11(3).
- [16] Kaur, J., & Saini, J. R. (2018). Automatic classification of Punjabi poetries using poetic features. *International Journal of Computational Intelligence Studies*, 7(2),124-137.<https://doi.org/10.1504/IJCISTUDIES.2018.094901>.
- [17] Pal, K., & Patel, B. V. (2020). Model for classification of poems in Hindi language based on ras. In *Smart Systems and IoT: Innovations in Computing:Proceeding of SSIC 2019* (pp. 655-661). Springer Singapore. https://doi.org/10.1007/978-981-13-8406-6_62
- [18] Lone, N. A., Giri, K. J., & Bashir, R. (2022). Machine Intelligence for Language Translation from Kashmiri to English. *Journal of Information & Knowledge Management*, 2250074.<https://doi.org/10.1142/S0219649222500745>.
- [19] Mir, T. A., & Lawaye, A. A. (2020, December). Word Sense Disambiguation For Kashmiri Language Using Supervised Machine Learning. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON)* (pp. 243-245), Indian Institute of Technology Patna, Patna, India. NLP Association of India (NLPAL).
- [20] Ahmad, A., & Syam, B. (2014). Kashmir part of speech tagger using CRF. *Computer Science*, 3(3), 3.
- [21] Rasool, N., Nabi, S., & Dar, Y. R. (2019). Machine Translation with UNL: A Study of Kashmiri. *Advances in Computer Science and Information Technology (ACSIT)*, Volume 6, Issue 2; April-June,2019, pp. 111-115.
- [22] Gilkar, S., Nahida, A. H., Nabi, S., & Farooq, V. (2014). Tagging: A Case Study of Kashmiri. *Interdisciplinary Journal of Linguistics*, 7, 248-261.
- [23] Aabid Rashid Wani, Er. Tabish Gulzar and Er. Mamoona Rashid (2015) "Kashmiri Speech Recognition System using Linear Predictive Coding and Artificial Neural Networks" *International Journal of Advanced Trends in Computer Applications (IJATCA)* Volume 2, Number 6,2015, pp. 20-23.
- [24] Thirumuru, R., Gurugubelli, K., & Vuppala, A. K. (2018). Automatic Detection of Palatalized Consonants in Kashmiri. *The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages*. <https://doi.org/10.21437/SLTU.2018-25>

- [25] Firdaus, E., Syahit, A. U., & Sukmawan, S. (2023, June). Poetry Writing Method Innovation Using "Atafora" Technique in Indonesian Language Learning. In Proceedings of the 2nd International Conference on Advances in Humanities, Education and Language, ICEL 2022, 07-08 November 2022, Malang, Indonesia.
- [26] Malik, N. A., Mustofa, A., & Munir, A. (2023). Unearthing the Figure of Speeches Used in the-R- Album by Rose to Deliver the Messages. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(2), 1787-1805.
- [27] Krisna, I. P. H. A., Jayantini, I. S. R., & Resen, I. W. (2023). The Types and Meaning of Paradox Found in the Poems of Rudyard Kipling's. Journal of Language and Applied Linguistics, 4(1), 11-16.
- [28] Maula, I. K., Sodiq, J., & Nugrahani, D. (2023, March). Analysis of figures of speech as used in songlyrics always remember us this way by ladygaga department of english education faculty of language and arts education universitas PGRI Semarang. In proceeding of english teaching, literature and linguistics (eternal) conference (Vol. 3, No. 1, pp. 124-134).
- [29] Wati, M. L. K., Supriyanto, R. T., & Baehaqie, I. (2023). The Comparative Figure of Speech in a Poetry Collection entitled Ibu Menanak Nasi hingga Matang Usia Kami by Emi Suy. Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia, 12(1), 43-52.
- [30] Naaz, K., & Singh, N. K. (2022). Design and development of computational tools for analyzing elements of Hindi poetry. IEEE Access, 10, 97733- 97747.
- [31] Mahdi, A. L. G. S., Mohammed, A. A. J., & Alsalmi, M. S. A. (2022). Writing Techniques in the English Love and Lyric Poems. Eurasian Journal of Research, Development and Innovation, 7, 18-21.
- [32] Nidi, V., Utami, N. M. V., & Maharani, P. D. (2022). An Analysis of Personification In The Some Selected Poems By Robert Frost. Journal of Humanities, Social Science, Public Administration and Management (HUSOCPUMENT), 2(2), 107-112.
- [33] Setiani, R. (2020). Denotative and connotative meaning used in writing poetry. Edukasi Lingua Sastra, 18(2), 85-92.
- [34] Hutaaruk, B. S. (2019). The Use of Figurative Languages on the Students' Poetry Semester V at FKIP Universitas HKBP Nommensen. Journal of English Language and Culture, 9(2).
- [35] Sayakhan, N. I. (2019). The use of personification and apostrophe as facilitators in teaching poetry. Journal of Language Studies, 1(4), 98-106.

Machine Learning Model for Automated Assessment of Short Subjective Answers

Zaira Hassan Amur¹, Yew Kwang Hooi², Hina Bhanbro³, Mairaj Nabi Bhatti⁴, Gul Muhammad Soomro⁵

Dept. Computer and Information Sciences, Universiti Teknologi PETRONAS, Perak, Malaysia^{1, 2, 3}

Dept. Information Technology, Shaheed Benazir Bhutto University, Nawabshah, Pakistan⁴

Dept. Information Technology, Tomas Bata University, Zlin, Czech Republic⁵

Abstract—Natural Language Processing (NLP) has recently gained significant attention; where, semantic similarity techniques are widely used in diverse applications, such as information retrieval, question-answering systems, and sentiment analysis. One promising area where NLP is being applied, is personalized learning, where assessment and adaptive tests are used to capture students' cognitive abilities. In this context, open-ended questions are commonly used in assessments due to their simplicity, but their effectiveness depends on the type of answer expected. To improve comprehension, it is essential to understand the underlying meaning of short text answers, which is challenging due to their length, lack of clarity, and structure. Researchers have proposed various approaches, including distributed semantics and vector space models. However, assessing short answers using these methods presents significant challenges, but machine learning methods, such as transformer models with multi-head attention, have emerged as advanced techniques for understanding and assessing the underlying meaning of answers. This paper proposes a transformer learning model that utilizes multi-head attention to identify and assess students' short answers to overcome these issues. Our approach improves the performance of assessing the assessments and outperforms current state-of-the-art techniques. We believe our model has the potential to revolutionize personalized learning and significantly contribute to improving student outcomes.

Keywords—Natural language processing; short text; answer assessment; BERT; semantic similarity

I. INTRODUCTION

Semantic similarity is a technique used to determine whether two separate texts have the same meaning. It is a crucial task in natural language processing (NLP) and can be applied to a range of downstream applications, such as text classification, summarization, and question-answering systems (QAS). In the early days of text similarity research, the emphasis was often on comparing lengthy texts, such as news articles, large corpora, and documents. Compared to lengthy writings, short texts have unique characteristics that pose challenges to traditional approaches for measuring similarity. First, short texts have a shorter form, which means that traditional approaches such as knowledge-based, and corpus-based which rely on examining common terms in two texts to determine similarity often lack statistical evidence to support them [1]. Second, short writings frequently use colloquial language and contain numerous typographical and grammatical errors. Third, due to the huge volume of short messages produced, they tend to be ambiguous and noisy [2]. Consequently, it is difficult to use traditional text similarity

methods for short texts. There are three main methods for calculating the similarity of short texts. The first method is word-level semantic-based, which looks at the words in the texts and finds pairs of similar words. It then calculates the similarity of the whole text based on the similarity of these word pairs. The second method is semantic modeling-based, which looks at the overall structure of the texts and compares the two models to see how similar they are. The third method is deep learning-based, which converts the short texts into "word embeddings" and calculates how close the words are to each other using cosine similarity [3]. Other approaches such as convolutional neural networks (CNN) and recurrent neural networks (RNN) can take a long time to train due to their sequential processing of information. However, most of the supervised work in NLP is done using a human-annotated corpus. This approach involves two steps: first, candidate phrases are extracted using a heuristic method, and then a classification model is trained to determine whether the phrase is from an answer or a sentence [4-5]. Other deep learning approaches such as pre-trained sentence transformers, like GPT-1, BERT, XLNet, Roberta, and ELECTRA, have been incredibly successful because they can learn a universal language representation from vast amounts of unlabeled text data. Moreover, the transformer learning model has led to a lot of progress in machine learning. It uses a sequence-to-sequence architecture and an attention mechanism to determine the significance of words in a sequence. This mechanism imitates the way humans read and think. Transformer-based models use a feature extraction technique that creates a vector for each word in the sequence based on its relevance to the other words. The aim of this study is to evaluate students' short subjective answers with the help of a multi-head attention transformer learning model based on the BERT language model, which is a promising method for improving the accuracy of student assessment. Student assessment is a crucial aspect of classroom instruction, involving evaluating students' knowledge, understanding, and skills to inform instruction and support student learning. Various forms of assessment, including quizzes, exams, projects, and presentations, serve to measure student progress and identify areas where additional support is needed [6-7]. By providing feedback to both students and teachers, effective assessment practices can help ensure that students are meeting learning goals and enable teachers to tailor instruction to meet the needs of individual students.

The key contributions of this study are:

- 1) To develop a BERT-based transformer learning model that utilizes multi-head attention to accurately identify and assess students' short answers in personalized learning.
- 2) To evaluate the effectiveness of the proposed transformer learning model in improving the accuracy of assessments compared to current state-of-the-art techniques.
- 3) To investigate the challenges of assessing short answers using machine learning methods, such as transformer models with multi-head attention, and propose solutions to overcome these issues.
- 4) To contribute to the field of NLP and personalized learning by developing an advanced technique for understanding and assessing the underlying meaning of short text answers.

The paper is structured as follows:

Section II provides a comprehensive review of the related literature, Section III elaborates on the proposed method, Section IV presents the results and corresponding discussions, and presents the implications and suggested solutions, and Section V concludes the paper.

II. RELATED WORK

The purpose of the literature review is to examine the various machine learning methods and techniques that are utilized in short text semantic similarity. In order to achieve this objective, we have meticulously analyzed 20 studies to provide a comprehensive overview of the related work.

The field of natural language processing (NLP) is experiencing an increasing use of deep learning techniques. Various attention-based neural network models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Bidirectional Encoder Representations from Transformers (BERT), have captured the attention of numerous researchers. In the domain of short text semantic similarity (STSS), Wang et al. [8] conducted a study in the field of short text semantic similarity. They utilized a Convolutional Neural Network (CNN) to classify short text and identify related words. The authors also used external knowledge and Jaro-Winkler similarity to understand the conceptual meaning of words and detect grammatical errors in sentences, respectively. The research conducted by Shih et al. [9] resulted in the development of a short answer grading system that utilizes a Convolutional Neural Network (CNN). Their CNN model operates as a text classifier to interpret Chinese language responses submitted by students. The authors' use of binary classification methods enabled the grading of answers as either correct or incorrect with relative accuracy. This system has the potential to revolutionize the grading process of short answer responses in the Chinese language.

Furthermore, the DE-CNN model introduced by Xu et al. [10] represents a significant advancement in the area of short-answer comprehension. By utilizing multiple embedding layers, the authors were able to gain a deeper understanding of the context and underlying concepts within short answers. The attention embedding layer further enabled the extraction of concept representations, enhancing the model's accuracy and

effectiveness. These findings have implications for the development of future deep-learning models designed for short-answer comprehension. Moreover, the CNN model proposed by Perera et al. [11] represents a significant contribution to the field of web-based question-answering systems. By focusing on factoid questions with short answers, the authors were able to develop a model that effectively identifies irrelevant answers. However, it is worth noting that the model's inability to answer the complete factoid question set highlights a need for continued research in this area. Future studies may consider building on this work to enhance the accuracy and effectiveness of web-based QAS models. The character-level CNN developed by Surya et al. [12] also represents a promising approach to short-answer comprehension. By relying solely on character-level data, the model can learn to identify key information without any prior knowledge of language or semantics. Nevertheless, the challenges faced by the authors in scoring short answers highlight the need for continued research in this area. Future studies may consider developing novel tools and strategies to enhance the effectiveness of short-answer scoring in the context of character-level CNNs. The approach proposed by Liu et al. [13] represents a novel approach to mining and comprehending global features from a short text. By combining the strengths of both CNN and LDA, the authors were able to develop a method that effectively captures both local and global features. This approach may have significant implications for natural language processing tasks, particularly in the context of short text comprehension. The SVMCNN model developed by Hu et al. [14] represents a promising approach to short-text classification. By combining the strengths of both CNN and SVM, the authors were able to develop a more robust model capable of accurately classifying short text data. Moreover, by training the model on the Twitter social platform using TensorFlow, the authors demonstrated the applicability of their method to real-world scenarios. This study may have important implications for the development of more effective short-text classification methods.

Moreover, the LSTM model proposed by Yao et al. [15] represents a novel approach to short text similarity calculation. By utilizing cosine similarity and backward propagation, the authors were able to develop a more accurate and robust model capable of measuring the similarity between short texts. This method may have important implications for various natural language processing tasks, such as information retrieval and question-answering systems. The approach taken by Zhou et al. [16] represents an innovative method for short-text classification using both RNN and CNN. By combining semantic features extracted from both types of neural networks, the authors were able to develop a more comprehensive and effective model for Chinese short-text classification. This study may have important implications for various natural language processing tasks, such as sentiment analysis and document classification. The study conducted by Hassan et al. [17] sheds light on the challenges involved in measuring similarity among short texts and proposes a potential solution using RNN and Tf-IDf vectors. The findings of this research could be beneficial for improving the performance of short text classification and similarity tasks in various fields, such as social media analytics, customer service, and content analysis.

The use of RNNs to generate vector representations of short text is a promising approach for improving the accuracy and efficiency of natural language processing tasks. The evaluation of the model on a standard benchmark dataset like DSTC provides a reliable way to assess its performance and compare it with other models. The research conducted by Lee et al. [18] contributes to the ongoing efforts to develop effective and scalable solutions for processing short text in various domains, such as social media, e-commerce, and customer service.

Furthermore, Mozafari et al. [19] introduced a BERT-based answer selection model (BAS) to capture both the syntactic and semantic information in short question-answer pairs. The model employs pre-trained BERT embeddings to encode the input text and utilizes a binary classification approach to predict whether a given answer is correct or not. The authors evaluated the performance of their proposed model on several benchmark datasets and achieved state-of-the-art results in short answer selection tasks. Wijaya et al. [20] leveraged the BERT model to devise an automated grading system for short answers in the Indonesian language. The study employed Cohen's Kappa to assess the inter-rater reliability among student answers, and the model demonstrated high accuracy in grading the short answers. These findings suggest that the proposed approach holds promise for implementation in educational contexts, where efficient grading mechanisms are essential.

Luo et al. [21] explored the use of the BERT model for grading short answers, similar to the study by Wijaya et al. [20]. However, they utilized a different dataset for training the model, which was the short answer scoring V2.0 dataset. In addition, the study used the regression task function to check the linearity between answers, and found that the BERT model achieved high accuracy in grading short answers. These results indicate the potential of the BERT model in improving the efficiency of grading mechanisms in educational settings. However, further research is needed to investigate the generalizability of these findings across different languages and domains. In the study, Alammary et al. [22] introduced the use of BERT models for short text classification in Arabic. The researchers explored different versions of BERT models and evaluated their effectiveness in classifying Arabic short texts. Furthermore, the study compared the performance of the Arabic BERT models with their English counterparts. This research has significant implications for natural language processing tasks in the Arabic language and can lead to the development of more effective and accurate models for Arabic text classification. Heidari et al. [23] developed a short answer grading system for Indonesian students using domain-independent subjects such as biology and geography. The study employed the BERT model to detect word embeddings from sentences and analyze the contextual information for improved grading accuracy. By integrating domain-specific knowledge into the model, the proposed approach demonstrated high accuracy in grading short answers, suggesting its potential use in educational contexts for efficient and reliable grading. Gaddipati et al. [24] highlighted the distinctions between transformer-based language models such as BERT, GPT, GPT2, and ELMO. Unlike ELMO and GPT, BERT utilizes a transformer mechanism and extracts contextual embeddings in

a bidirectional manner. The model is trained on large-scale datasets such as Book Corpus and Wikipedia, which consist of 800M and 2500M words, respectively. Overall, the study sheds light on the unique features and capabilities of these advanced language models. Furthermore, Zhu et al. [25] developed a BERT-based framework for grading short answers, which incorporated CNN and capsule networks, as well as a triple-hot loss strategy to encode key sentences. The approach was tested on a dataset of student short-answer responses and yielded superior results compared to other state-of-the-art methods. These findings indicate that the proposed framework has the potential to significantly improve the accuracy and efficiency of grading mechanisms in educational settings. In addition to the classification of ASAG systems, Burrow et al. [26] also identified several limitations in these systems. One of the main limitations identified was the inability of existing ASAG systems to handle complex or open-ended questions. Another limitation was the reliance of ASAG systems on pre-defined rubrics, which can limit the flexibility of the grading process. The authors suggested that future research should focus on addressing these limitations and developing more sophisticated ASAG systems that can handle a wider range of questions and provide more accurate and flexible grading mechanisms.

On the other hand, Mohler et al. [27] proposed a different approach for grading short answers using lexical semantic similarity. The authors argue that deep learning techniques, such as the ones used in BERT and other transformer-based models, may not be the most effective method for grading short answers because they rely on large amounts of training data. Instead, Mohler et al. utilized a method based on semantic similarity to assess the quality of short answers. By comparing the semantic features of the correct answer and the student's answer, the system was able to assign a score that reflected the level of correctness. This approach may be particularly useful for assessing short answers in domains where training data is limited, and where deep learning models may not be effective. Ye et al. [28] leveraged the BERT model to generate context-sensitive representations and combined it with the GCN model to classify short text. The study demonstrated that the proposed approach achieved high accuracy in classifying short text, indicating its potential application in various natural language processing tasks. By incorporating both contextual and graph-based information, the proposed method may provide a more comprehensive understanding of the meaning of the short text.

III. METHODOLOGY

A. Data Collection

This study utilizes the computer science dataset, developed by Mohler et al. [27], which comprises 2443 student answers and 87 questions from 12 assignments in the field of computer science. The dataset includes both, the questions and reference answers, as well as the student responses, and was designed to evaluate the effectiveness of models in grading student answers by comparing them to the evaluator's desired answer. The dataset has been graded by human evaluators who are experts in the field of computer science, and the grading scale ranges from 0 (not correct) to 5 (totally correct). Table I provides a detailed overview of the dataset used in this study. It contains a total of 87 questions and 2442 student responses, which are

distributed across 12 different assignments. The number of questions in each assignment varies. The grading of each student response is done by two human graders, and the average of their scores is used as the standard score for each response.

TABLE I. MOHLER DATASET FOR THE ASSESSMENT

Institute	University of North Texas
Class/domain	Introductory computer science class
Course	Data structure
Assignments	12 assignments
Questions	87
Answers	2442
Score by human evaluators	0-5
Data collection	WebCT online learning environment
Type of questions	Open-ended

Additionally, it is worth noting that the dataset used in this study exhibits a bias toward correct answers. The dataset comprises both very short and very long answers. However, to ensure a fair evaluation, we have only included answers containing 10-20 words from the test dataset, which is considered an ideal size for short answers. Fig. 1(a) and Fig. 1(b) present the score assigned by human evaluators which is inconsistent, and Fig. 1(b) illustrates the biased nature of the dataset toward correct answers.

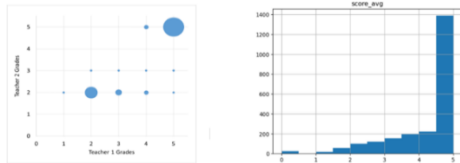


Fig. 1. (a) Distribution of grades and (b) Histogram of scores assigned by human evaluators [29].

Table II illustrates the sample example from the dataset which contains questions, teacher answers, student answers, and scores assigned by teachers.

TABLE II. ILLUSTRATES THE SAMPLE EXAMPLE FROM THE DATASET

	Sample of questions, teacher answers, and student answers	Grades
Question. Teacher answer.	What is a variable? A location in memory that can store a value.	
Student answer:	A block of memory that holds a specific type of data.	5,5
Student answer:	A pointer to a location in memory.	3,5
Question. Teacher answer.	What is a pointer? A variable that contains the address in memory of another variable.	
Student answer:	A pointer holds a memory location.	5,4
Student answer:	Is a reference call to the place in memory where the object is stored.	3,4

B. Pre-processing

Before training and testing our machine learning model, we performed several pre-processing steps on the dataset to ensure that the text data was in a clean and normalized format. One issue we encountered was that many of the student answers contained spelling errors and unnecessary punctuation, which can introduce noise and sparsity into the dataset and negatively impact the performance of our model. To address this issue, we implemented techniques such as spell-check and punctuation removal to clean and normalize the text data. These steps involved identifying and correcting misspelled words, and removing unnecessary punctuation marks that may interfere with our analysis. Additionally, we also performed other pre-processing steps as mentioned in Fig. 2 such as removing stop words and converting text to lowercase to further enhance the quality and consistency of the data. By carefully pre-processing the dataset, we were able to significantly improve the accuracy and effectiveness of our machine-learning model, and ultimately generate more reliable and informative results. The following algorithm 1 mentions the pre-processing steps applied to the dataset (see Fig. 2).

Pseudocode of Pre-processing Dataset	
Input:	<ul style="list-style-type: none"> A dataset with text entries that contain spelling errors and or punctuation.
Output:	<ul style="list-style-type: none"> A cleaned dataset with the same text entries, but with all spelling errors and punctuation removed.
Steps:	<ol style="list-style-type: none"> Load the dataset into a list. Define a function to remove punctuation from a given text. Define a function to remove spelling errors from a given text. Loop through each entry in the dataset and apply the following cleaning steps: <ol style="list-style-type: none"> Remove punctuation from the entry using the punctuation removal function. Remove spelling errors from the entry using the spelling error removal function. Append the cleaned entry to a new list of cleaned data. Return the new list of cleaned data.

Fig. 2. Cleaning of the dataset.

C. BERT-Multi-Head Attention Model

BERT is a transformer-based model that utilizes bidirectional processing and attention mechanism to understand language. Several versions of the BERT model such as Roberta, KeyBERT, M-BERT, and SBERT have been introduced. We used the BERTbase-uncased model as it has shown the best performance on NLP tasks. This model can encode various languages but utilizes the default English vocabulary. The architecture of the BERT model with tokens is shown in Fig. 3, where E1-EN generates the input tokens, and T1-TN are output tokens that categorize phrases using binary representations and deliver them to the C-label. BERT employs a masked language modeling technique that predicts incoming words based on the surrounding context. This technique changes 15% of the words in a sentence presented in Fig. 4 where 80% are converted to "mask" tokens, 10% to random words, and 10% to their previous representations. BERT evaluates the accuracy of its predictions and fine-tunes accordingly. Compared to an implementation of BERT that operates without masking, BERT coverage is slower but reaches a higher threshold. The next sentence prediction (NSP) examines whether the two sentences are connected logically, providing contextual information for both sentences.

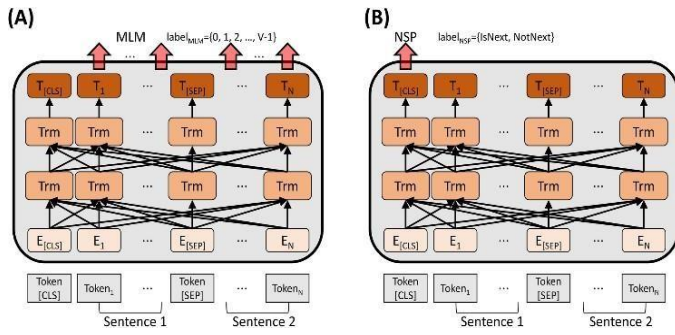


Fig. 3. BERT language model.

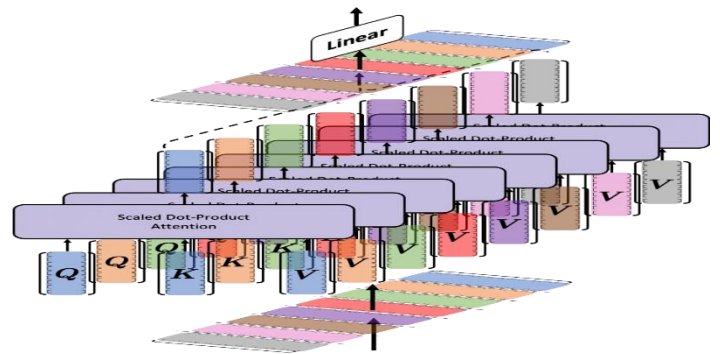


Fig. 5. Multi-head attention.

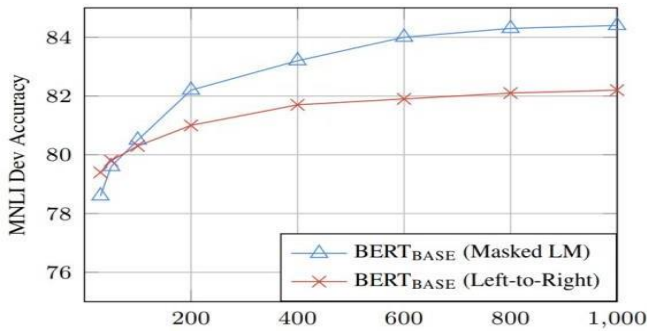


Fig. 4. BERT model with and without masked language modeling.

It uses a transformer-based architecture to process text data and learn contextual representations. One of the key components of transformer architecture is multi-head attention. Multi-head attention is a type of attention mechanism that allows the model to attend to different parts of the input sequence simultaneously. In the case of BERT, multi-head attention is used in both the encoder and decoder components of the transformer architecture. The BERT multi-head attention model consists of three components: query, key, and value matrices as shown in Fig. 5. These matrices are used to compute the attention scores, which determine how much attention each token in the input sequence should receive. The query matrix represents the current token that is being processed, while the key and value matrices represent all the other tokens in the sequence. The multi-head attention mechanism in BERT involves splitting the query, key, and value matrices into multiple heads. Each head has its own set of parameters and is trained to attend to a different part of the input sequence. This allows the model to capture different aspects of the input sequence and learn more complex relationships between the tokens. The output of the multi-head attention mechanism is computed as the weighted sum of the values, where the weights are determined by the attention scores. The attention scores are computed by taking the dot product of the query matrix and the key matrix and then applying a SoftMax and linear function to normalize the scores. The resulting attention vector is then multiplied by the value matrix to obtain the output. The BERT multi-head attention model also includes a layer normalization step after the output is computed. Layer normalization ensures that the output has a mean of zero and a standard deviation of one, which helps to improve the stability and performance of the model.

The implementation method utilized the BERT multi-head attention model, a monolingual model solely evaluated on the English dataset. The process involved setting up sentence transformers on a dataset and fine-tuning the model through a question-answer task. The results were evaluated using statistical approaches. Fine-tuning the model ensured that it could accurately understand the context of the given task, making it suitable for various natural language processing applications. Additionally, the implementation method applied the attention mechanism to highlight the answers that best matched the teacher's answer. This approach enabled the model to provide more accurate responses and perform better in question-answering tasks. Fig. 6 presents a visual representation of the implementation process. By following this process, the model can be optimized for specific tasks, resulting in better performance. It is worth noting that this implementation method is limited to the English language, as BERT is a monolingual model. However, there are other models available that support multiple languages. Overall, the BERT multi-head attention model is a powerful method for natural language processing tasks and has been widely adopted in various applications, including chatbots, sentiment analysis, and language translation.

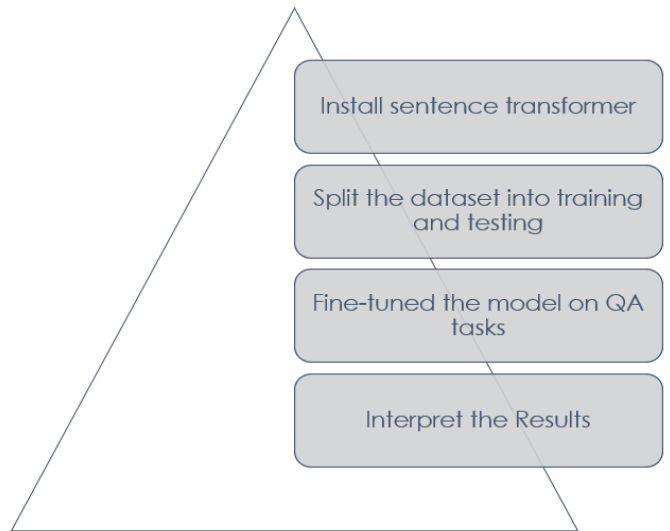


Fig. 6. Implementation details.

D. Training and Testing

To evaluate the performance of our approach, we utilized the BERT multi-head attention model. We randomly split the Mohler dataset into 70% training and 30% testing data, ensuring that the split was representative of the entire dataset. We trained the model for 1000 iterations, using different training and testing data for each iteration to improve the generalization of the results. The cosine similarity feature was trained using isotonic, linear, and non-linear (ridge) regression models, and the performance was compared to previously established models such as Mohler et al. [27]. Following training, we evaluated the model by testing it on the unseen test data. The similarity scores of the test data were input through the trained regression model, resulting in predicted grades that were compared to the desired scores. We calculated the Root Mean Square Error (RMSE) and Pearson correlation to evaluate the model's performance. Our utilization of the BERT multi-head attention model allowed us to effectively analyze and evaluate the performance of our approach on the Mohler dataset. The results of this study could have significant implications for future natural language processing applications, particularly those that require accurate grading of written responses.

IV. RESULTS

A. Feature Extraction

In the feature extraction, we use the pre-trained embeddings from each transfer learning model and assign them to the tokens of every word in all the answers. To create answer embeddings, we use the Sum of Word Embeddings (SOWE) method, as shown in Eq. (1). Here, a_{ij} denotes the j -th answer vector of question q_i , and w_k represents the vector of the k -th word in the answer a_{ij} . By applying this method, we obtain a single vector that represents each answer in a high-dimensional hypothesis space. The resulting sentence embeddings have the same size as the word embeddings. This approach allows us to capture the semantic and syntactic properties of each answer and create a compact representation of it.

$$a_{ij} = \sum_{k=1}^{n_j} w_k \quad (1)$$

In this equation, " a_{ij} " represents the vector of the j th answer of the question " q_i ", " w_k " represents the vector of the k th word in the answer " a_{ij} ", and " n_j " represents the number of words in the answer " a_{ij} ". The equation calculates the sum of the word embeddings for each word in the answer to create a single vector representing the entire answer. To create a Question-Answering model using BERT, the tokenizer utilizes two special tokens, namely [CLS] and [SEP]. These tokens serve the purpose of encoding the sentence sequence. The [CLS] token is a classification token, whereas the [SEP] token separates the Key and response answer, as exemplified below. The sequence of sentences is then passed as a token input to the BERT model for training [32,35]. The model generates high-dimensional embeddings for input tokens, which are then used to predict the grades within a specified range. An example of BERT embeddings is given below.

Question: What is a variable??

Key Answer: A location in memory that can store a value.

Student Answer: a block of memory that holds a specific type of data.

[CLS] and [SEP]: [CLS] a location in memory that can store a value [SEP] A block of memory that holds a specific type of data

Token ids of both responses:

[101, 10408, 1996, 9896, 1998, 5468, 1995, 3558, 2770, 2051, 1012, 102, 270, 2019, 9896, 2006, 103, 3563, 102]

Example of Response pair and Token Id's

Additionally, the multi-head attention mechanism is utilized to visualize the relationships between words. The lines that are darker in color indicate a closer relationship between words at layers 1,2,3,4.

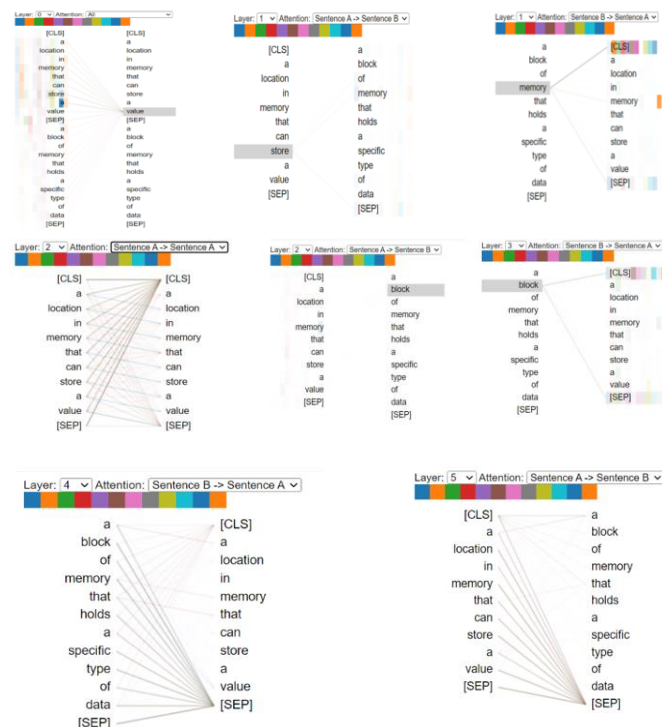


Fig. 7. BERT-based multi-head attention model for layers 0, 1, 2, 3, 4, 5

The findings obtained from the BERT-based multi-head attention model are extremely encouraging (see Fig. 7). Layer 0 has highlighted the importance of the [CLS] and [SEP] tokens, as they effectively emphasize the embeddings from the text. Layer 1 has shown a strong correlation between the words "store" and "memory," indicating that they are related. In layers 2, 3, and 4, words such as "block," "value," "memory," "type," and "hold" also share strong embeddings, indicating their interconnectedness. Moreover, the relationship between student and teacher responses has also been established, as their embeddings show strong correlations. Layer 0 also performed self-attention with multi-heads to determine the relationships between different parts of the answer itself. The multi-head attention module utilized teacher-to-teacher, teacher-to-student, and student-to-student attention to determine the strong impact

of different words. This approach has proved to be highly effective in identifying the key components of the response and the underlying relationships between them. The results obtained through this model have significant implications for natural language processing, particularly in the field of language understanding and interpretation. The multi-head attention mechanism can be used to improve the accuracy of machine translation, question-answering systems, and text classification algorithms, among others. Overall, the BERT-based multi-head attention model has demonstrated its ability to capture complex relationships between different components of natural language text. This approach has the potential to revolutionize the field of natural language processing and enable more accurate and efficient analysis of text data.

B. Automated Scoring

To evaluate student answers, we employed text similarity techniques that match the embeddings of teacher answers to student responses. We used Python programming with Spyder IDE to implement this process. After extracting the embeddings using the BERT-multi-head attention model, we used four similarity methods to obtain a similarity value, which we used as a scoring rubric. These four methods are the longest common subsequence (LCS), cosine coefficient (SC), Jaccard coefficient (JC), and Dice coefficient (DC). Additionally, we chose these methods as they are commonly used in Natural Language Processing (NLP) and provide reliable results for text similarity comparisons.

$$Score = \frac{(Score(sim)*answer\ score)+(Score(keymatch)*answer\ score)}{2} \quad (2)$$

Example to calculate the similarity of teacher and student answers:

Teacher answer: a location in memory that can store a value.

Student answer: a block of memory that holds a specific type of data.

$$Sim_{lcs} = \frac{2*40}{58*40} = 0,81633 \quad sim(cosine) = \sqrt[6]{9*6} = 0,81650 \quad score(keymatch) = 6/9 = 0,66667$$

$$Sim_{jaccard} = \frac{6}{9} = 0,66667 \quad Simdice = \frac{2*6}{9+6} = 0,80000$$

To determine the score for each student response, we derived a scoring rubric by averaging the similarity values obtained through two different methods: the String-based method and the keyword-matching technique. The String-based method involves comparing the strings in the form of embedding values of the teacher's and student's answers to identify common sub-sequences and measure their similarity. On the other hand, the keyword matching technique involves identifying the presence of specific keywords in the student's answer that are expected based on the question or prompt.

In the previous example, we obtained a similarity score of 0.66667. However, it should be noted that we have multiple references available, and we select the highest similarity score among them. In this case, the highest similarity score is 0.85714.

$$Score = \frac{(0.81633*4)+(0.85714*4)}{2} = 3.34695$$

Therefore, we consider this score as the final similarity score for the response in question. This approach helps to ensure that the students receive a fair and accurate evaluation, as we consider all available references and select the most appropriate one.

C. Comparative Evaluation

Our model's performance was evaluated using RMSE and Pearson correlation scores, and we conducted a comparative analysis of our model's performance with various pre-trained models, such as ELMO, GPT, and GPT2, as reported by Gaddipati et al. [24], on the Mohler dataset. We further compared our model's performance with other approaches and showed that the BERT-multi-head attention method outperformed other techniques in terms of effectiveness. Table III displays the Root Mean Square Error (RMSE) and Pearson correlation (ρ) results of various Pre-trained transfer learning models on the Mohler Dataset.

Table IV compares the performance of different models and approaches on the Mohler dataset. The models are evaluated based on their RMSE (Root Mean Square Error) and Pearson correlation scores. The results for the BOW (Bag of Words) approach with SVMRank, achieve an RMSE of 1.042 and a Pearson correlation score of 0.480. The results for the tf-idf approach with SVR, which performs slightly better than the BOW approach with an RMSE of 1.022 and a Pearson correlation score of 0.327. The results for the tf-idf approach with LR (Logistic Regression) and SIM (Semantic Information), which outperforms the previous two approaches with an RMSE of 0.887 and a Pearson correlation score of 0.592. Furthermore, the results for three different word embedding models - Word2Vec, GloVe, and FastText; all of these models use SOWE (Sum of Word Embeddings) and Verb phrases features, and they achieve RMSE values ranging from 1.023 to 1.036 and Pearson correlation scores ranging from 0.425 to 0.465. The results for deep learning models - ELMo, GPT-2, and Roberta; ELMo uses a 5-layer BiLSTM (Bidirectional Long Short-Term Memory) with max-pooling and achieves an RMSE of 0.875 and a Pearson correlation score of 0.655. GPT-2 uses a 12-layer Transformer and achieves an RMSE of 0.911 and a Pearson correlation score of 0.610. Roberta uses a 24-layer Transformer and achieves the best performance among all the models with an RMSE of 0.851 and a Pearson correlation score of 0.692. Results of our BERT-based model (our approach) with Multihead Attention as the feature has RMSE value as 1.990, which means that on average, our model's predictions deviate from the actual values by 1.990 points. The Pearson correlation coefficient is 0.773, which indicates a strong positive correlation between our model's predicted scores and the actual scores. Overall, the RMSE value of 1.990 is higher than the RMSE value of the RoBERTa model (0.851), which indicates that our model has a higher prediction error than RoBERTa. However, the Pearson correlation coefficient of our model (0.773) is higher than that of RoBERTa (0.692), indicating that our model's predicted scores are more strongly correlated with the actual scores than RoBERTa's predictions.

TABLE III. A SUMMARY OF THE RESULTS OBTAINED BY VARIOUS APPROACHES ON THE MOHLER DATASET IS PRESENTED

Model	Isotonic regression		Linear regression		Ridge regression	
	RMSE	P	RMSE	p	RMSE	P
ELMO [24]	0.978	0.485	0.995	0.451	0.996	0.449
GPT [24]	1.082	0.248	1.088	0.222	1.089	0.217
GPT2 [24]	1.065	0.311	1.077	1.075	1.079	0.269
BERT-multi-head attention Our model	1.089	0.456	1.990	0.773	1.536	0.876

TABLE IV. COMPARISON WITH OTHER METHODS

Model	Features	RMSE	Pearson Correlation
BOW [29]	SVMRank	1.042	0.480
Tf-idf [30]	SVR	1.022	0.327
Tf-idf [31]	LR+SIM	0.887	0.592
Word2Vec[33]	SOWE + Verb Phrases	1.025	0.458
Glove [34]	SOWE + Verb Phrases	1.036	0.425
FastText[35]	SOWE+Verb Phrases	1.023	0.465
ELMO [24]	5-layer BiLSTM +max-pooling	0.875	0.655
GPT-2 [24]	12-layer transformer	0.911	0.610
Roberta[36]	24-layer transformer	0.851	0.692
BERT (our approach)	Multihead Attention	1.990	0.773

D. Model Implications

One of the primary challenges is the limited availability of training data for short answers. Short answers are usually context-dependent and diverse in nature, making it difficult to generate large amounts of high-quality training data. Additionally, there is often ambiguity and variation in short answers, which makes it challenging for machine learning models to accurately evaluate them. Another challenge is the need for efficient methods to encode short answers and generate embeddings that can be used for similarity matching. Transformer models with multi-head attention have shown promise in this regard, but there is a need for further research to optimize their performance for short answer evaluation. Furthermore, there is a need for developing robust methods to handle outliers, exceptions, and edge cases that are often encountered in short answer assessment. This requires careful consideration of the characteristics of short answers and the

design of models and algorithms that can handle such situations effectively [36-40]. To overcome these challenges, potential solutions include utilizing data augmentation techniques to generate more diverse training data, developing novel algorithms and models specifically tailored for short answer assessment, and leveraging domain-specific knowledge and expertise to enhance the performance of machine learning models. Additionally, the use of ensemble methods and human-in-the-loop approaches may improve the accuracy and reliability of short answer evaluation.

V. CONCLUSION AND FUTURE WORK

This study aimed to assess short subjective answers using a BERT-based multi-head attention model and string-based methods such as cosine co-efficient, longest common subsequence, Dice coefficient, and Jaccard coefficient to score the answers. Additionally, we compared the performance of the BERT multi-head attention model with former approaches using isotonic, linear, and ridge regression. The findings suggest that the BERT multi-head attention model outperforms other approaches, indicating its effectiveness in understanding and assessing the underlying meaning of short answers. Our study highlights the potential of machine learning methods in improving the efficiency of personalized learning, particularly in the assessment of open-ended questions. Overall, this study contributes to the growing body of research on NLP techniques and their applications in the education domain. Further research can explore the generalizability of our proposed model in different educational settings and subject domains.

ACKNOWLEDGMENT

Appreciation goes to the Pre-commercialization-External: YUTP-PRG Cycle 2022 (015PBC-005)

REFERENCES

- [1] B. Agarwal, H. Ramampiaro, H. Langseth, M. J. I. P. Ruocco, and Management, "A deep network model for paraphrase detection in short text messages," vol. 54, no. 6, pp. 922-937, 2018.
- [2] Z. H. Amur, Y. Kwang Hooi, H. Bhanbhro, K. Dahri, and G. M. J. A. S. Soomro, "Short-Text Semantic Similarity (STSS): Techniques, Challenges and Future Perspectives," vol. 13, no. 6, p. 3911, 2023.
- [3] Z. H. Amur, Y. K. Hooi, and G. M. Soomro, "Automatic Short Answer Grading (ASAG) using Attention-Based Deep Learning MODEL," in 2022 International Conference on Digital Transformation and Intelligence (ICDI), 2022.
- [4] H. Bhanbhro, Y. K. Hooi, and Z. Hassan, "Modern Approaches towards Object Detection of Complex Engineering Drawings," in 2022 International Conference on Digital Transformation and Intelligence (ICDI), 2022.
- [5] N. Carlini et al., "Extracting training data from large language models," in 30th USENIX Security Symposium (USENIX Security 21), 2021.
- [6] H. Bhanbhro, Y. Kwang Hooi, W. Kusakunniran, and Z. H. J. A. S. Amur, "A Symbol Recognition System for Single-Line Diagrams Developed Using a Deep-Learning Approach," vol. 13, no. 15, p. 8816, 2023.
- [7] L. B. Galhardi and J. D. Brancher, "Machine learning approach for automatic short answer grading: A systematic review," in Ibero-american conference on artificial intelligence, 2018.
- [8] H. Wang, K. Tian, Z. Wu, and L. J. I. J. o. C. I. S. Wang, "A short text classification method based on convolutional neural

- network and semantic extension," vol. 14, no. 1, pp. 367-375, 2021.
- [9] S.-H. Wu and C.-Y. Yeh, "A Short Answer Grading System in Chinese by CNN," in 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), 2019.
- [10] J. Xu et al., "Incorporating context-relevant concepts into convolutional neural networks for short text classification," vol. 386, pp. 42-53, 2020.
- [11] N. Perera, C. Priyankara, and D. Jayasekara, "Identifying Irrelevant Answers in Web Based Question Answering Systems," in 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), 2020.
- [12] K. Surya, E. Gayakwad, and M. J. I. J. R. T. E. Nallakaruppan, "Deep learning for short answer scoring," vol. 7, no. 6, pp. 1712-1715, 2019.
- [13] J. Liu, H. Ma, X. Xie, and J. J. E. Cheng, "Short Text Classification for Faults Information of Secondary Equipment Based on Convolutional Neural Networks," vol. 15, no. 7, p. 2400, 2022.
- [14] Y. Hu, J. Ding, Z. Dou, H. J. C. I. Chang, and Neuroscience, "Short-text classification detector: a bert-based mental approach," vol. 2022, 2022.
- [15] L. Yao, Z. Pan, and H. J. I. A. Ning, "Unlabeled short text similarity with LSTM encoder," vol. 7, pp. 3430-3437, 2018.
- [16] Y. Zhou, B. Xu, J. Xu, L. Yang, and C. Li, "Compositional recurrent neural networks for chinese short text classification," in 2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI), 2016.
- [17] Z. H. Amur and Y. K. J. I. S. L. Hooi, "State-of-the-Art: Assessing Semantic Similarity in Automated Short-Answer Grading Systems," vol. 11, pp. 1851-1858, 2022.
- [18] J. Y. Lee and F. J. a. p. a. Dernoncourt, "Sequential short-text classification with recurrent and convolutional neural networks," 2016.
- [19] J. Mozafari, A. Fatemi, and M. A. J. a. p. a. Nematbakhsh, "BAS: an answer selection method using BERT language model," 2019.
- [20] M. Wijaya, "Automatic Short Answer Grading System in Indonesian Language Using BERT Machine Learning," vol. 35, no. 6, pp. 503-509, 2021.
- [21] J. Luo, "Automatic Short Answer Grading Using Deep Learning," Illinois State University, 2021.
- [22] A. S. J. A. S. Alammary, "BERT Models for Arabic Text Classification: A Systematic Review," vol. 12, no. 11, p. 5720, 2022.
- [23] M. Heidari, J. H. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in 2020 International Conference on Data Mining Workshops (ICDMW), 2020, pp. 480-487: IEEE.
- [24] S. K. Gaddipati, D. Nair, and P. Plöger, "Comparative evaluation of pretrained transfer learning models on automatic short answer grading," 2020.
- [25] X. Zhu, H. Wu, and L. J. I. T. o. L. T. Zhang, "Automatic Short-Answer Grading via BERT-Based Deep Neural Networks," vol. 15, no. 3, pp. 364-375, 2022.
- [26] S. Burrows, I. Gurevych, and B. J. I. J. Stein, "The eras and trends of automatic short answer grading," vol. 25, no. 1, pp. 60-117, 2015.
- [27] M. Mohler, R. Bunescu, and R. Mihalcea, "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments," in Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies, 2011, pp. 752-762.
- [28] Z. Ye, G. Jiang, Y. Liu, Z. Li, and J. Yuan, "Document and word representations generated by graph convolutional network and bert for short text classification," in ECAI 2020: IOS Press, 2020, pp. 2275-2281.
- [29] N. Süzen, A. N. Gorban, J. Levesley, and E. M. J. P. c. s. Mirkes, "Automatic short answer grading and feedback using text mining methods," vol. 169, pp. 726-743, 2020.
- [30] R. Mihalcea and P. Tarau, "Textrank: Bringing order into text," in Proceedings of the 2004 conference on empirical methods in natural language processing, 2004, pp. 404-411.
- [31] S. Jimenez, S.-P. Cucerzan, F. A. Gonzalez, A. Gelbukh, G. J. J. o. I. Dueñas, and F. Systems, "BM25-CTF: Improving TF and IDF factors in BM25 by using collection term frequencies," vol. 34, no. 5, pp. 2887-2899, 2018.
- [32] K. W. Church, "Word2Vec," vol. 23, no. 1, pp. 155-162, 2017.
- [33] H. Al-Bataineh, W. Farhan, A. Mustafa, H. Seelawi, and H. T. Al-Natshah, "Deep contextualized pairwise semantic similarity for arabic language questions," in 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019.
- [34] D. Cer et al., "Universal sentence encoder," 2018.
- [35] W. Hua, Z. Wang, H. Wang, K. Zheng, and X. Zhou, "Short text understanding through lexical-semantic analysis," in 2015 IEEE 31st international conference on data engineering, 2015, pp. 495-506: IEEE.
- [36] A. Hassan and A. Mahmood, "Deep learning approach for sentiment analysis of short texts," in 2017 3rd international conference on control, automation and robotics (ICCAR), 2017.
- [37] Z. H. Amur, Y. K. Hooi, G. M. Soomro, H. Bhanbhro, S. Karyem, and N. J. A. S. Sohu, "Unlocking the Potential of Keyword Extraction: The Need for Access to High-Quality Datasets," vol. 13, no. 12, p. 7228, 2023.
- [38] M. A. Memon, Z. Hassan, K. Dahri, A. Shaikh, and M. A. J. I. Nizamani, "Aspect Oriented UML to ECORE Model Transformation," vol. 11, no. 3, 2019.
- [39] Z. Hassan, Z. Bhatti, K. J. U. o. S. J. o. I. Dahri, and C. Technology, "A conceptual framework development of the social media learning for undergraduate students of University of Sindh," vol. 3, no. 4, pp. 178-184, 2019.
- [40] A. R. Gilal, A. Waqas, B. A. Talpur, R. A. Abro, J. Jaafar, and Z. H. Amur, "Question Guru: An Automated Multiple-Choice Question Generation System," in International Conference on Emerging Technologies and Intelligent Systems, 2022.

Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach

Helmi Imaduddin¹, Fiddin Yusufida A'la², Yusuf Sulisty Nugroho³

Department of Informatics, Universitas Muhammadiyah Surakarta, Indonesia, Surakarta, Indonesia^{1,3}

Department of Informatics Engineering, Universitas Sebelas Maret, Surakarta, Indonesia²

Abstract—The rapid growth of application development has made applications an integral part of people's lives, offering solutions to societal problems. Health service applications have gained popularity due to their convenience in accessing information on diseases, health, and medicine. However, many of these applications disappoint users with limited features, slow response times, and usability challenges. Therefore, this research focuses on developing a sentiment analysis system to assess user satisfaction with health service applications. The study aims to create a sentiment analysis model using reviews from health service applications on the Google Play Store, including Halodoc, Alodokter, and klikdokter. The dataset comprises 9,310 reviews, with 4,950 positive and 4,360 negative reviews. The IndoBERT pre-training method, a transfer learning model, is employed for sentiment analysis, leveraging its superior context representation. The study achieves impressive results with an accuracy score of 96%, precision of 95%, recall of 96%, and an F1-score of 95%. These findings underscore the significance of sentiment analysis in evaluating user satisfaction with health service applications. By utilizing the IndoBERT pre-training method, this research provides valuable insights into the strengths and weaknesses of health service applications on the Google Play Store, contributing to the enhancement of user experiences.

Keywords—Application; healthcare; IndoBERT; sentiment analysis

I. INTRODUCTION

In the present age of digital advancements, a myriad of applications has emerged to cater to diverse facets of human life, spanning across desktops, tablets, and smartphones. This surge in demand has created lucrative business prospects, resulting in the proliferation of mobile applications aimed at resolving everyday challenges [1]. Regrettably, not all applications boast commendable features and functionalities, including the healthcare applications readily available on the Google Play Store. Consequently, there arises a pressing need for a system capable of comprehensively analyzing application reviews to enhance their overall performance. While opinions and ratings serve as primary means for gathering feedback on an app's usability, ratings alone may not consistently provide reliable insights [2]. Furthermore, ratings fail to offer a comprehensive understanding to improve the user experience aspect. Thus, the examination of customer reviews becomes crucial for gaining deeper insights and understanding [3]. User experience entails the intricate narrative surrounding a user's interaction with the app, while opinions delve into their underlying thoughts and emotions. Users possess the freedom to express their evaluations in various textual forms, resulting

in a less structured review dataset, which in turn poses greater challenges in handling and analysis [4].

Sentiment analysis, commonly referred to as opinion mining, is a technique that aims to classify user sentiment based on polarity [5]. It encompasses a wide array of objectives, methodologies, and types of analytics. In the domain of sentiment analysis, three main methodologies are employed: machine learning (ML), hybrid learning, and lexicon-based approaches [6]. Among these, supervised learning emerges as the most popular and widely utilized ML approach. This methodology involves training the model using labeled data to predict outputs, while also incorporating additional unlabeled inputs for enhanced performance [7].

In the context of sentiment analysis, it is important to acknowledge that certain languages, such as English and Chinese, benefit from being considered high-resource languages, as they have readily available datasets accessible to the academic community. However, the majority of languages face challenges due to limited data collection and a lack of published research, including Indonesian [8]. Previous research on sentiment analysis of Indonesian text has explored the efficacy of machine learning models like Support Vector Machine (SVM) and Naïve Bayes, demonstrating their effectiveness in addressing this issue [9]. Nevertheless, the integration of pre-training using language models has emerged as a promising approach across various natural language processing tasks [10], [11]. One significant drawback of conventional language models is their unidirectional nature, which imposes limitations on the available architecture for pre-training. To overcome this limitation, a novel technique called Bidirectional Encoder Representations from Transformers (BERT) has been proposed to enhance the fine-tuning-based approach [12].

A number of previous studies have explored sentiment analysis in the context of Indonesian language, employing various approaches ranging from traditional machine learning classifiers to deep learning-based algorithms such as IndoBERT. For instance, sentiment analysis using random forest algorithms demonstrated promising results, achieving an average out-of-bag (OOB) score of 0.829 [13]. Similarly, research focused on emoticons and emoticon categories utilized classification-based machine learning algorithms like naïve Bayes and support vector machines [14]. Sarcasm data classification was also conducted using random forest classifiers, naïve Bayes, and support vector machines [14]. Furthermore, Word2Vec was employed as an alternative to hand-crafted features for sentiment analysis of hotel reviews in

Indonesian, with the conclusion that optimal accuracy can be achieved by simultaneously increasing vector dimensions and the amount of data [15]. IndoBERT, which outperforms both multi-Indonesian lingual BERT and Bert, has demonstrated superior data processing capabilities. The research involved data collection, data preprocessing, and fine-tuning of IndoBERT. Hoax detection classification was completed using pre-trained BERT models, with multilingual BERT for general purposes and IndoBERT specifically tailored for Indonesian. The fine-tuned IndoBERT model, trained on an Indonesian monolingual corpus, exhibited enhanced performance compared to the original BERT and improved multilingualism [16].

However, performing this analysis manually is quite difficult; hence we propose performing Indonesian language analysis using the IndoBERT algorithm, which was built exclusively to evaluate Indonesian language material. The primary objective of this research is to develop a sentiment analysis system for healthcare application reviews on the Google Play Store using the IndoBERT approach. Additionally, the system aims to assist users in selecting health service applications that offer optimal functionality and facilities. The proposed methodology involves leveraging IndoBERT as a pre-training model, renowned for its effectiveness in processing Indonesian language data. The data utilized in this research is sourced from the Google Play Store, making it a novel and previously unexplored area of investigation.

II. METHODOLOGY

The research conducted in this study encompasses several stages. Firstly, review data was collected by scraping the Google Play website, followed by manual labeling of the obtained data. The labeled data was then preprocessed to ensure cleanliness and suitability for classification. The dataset was subsequently divided into three parts: training data, validation data, and testing data, with a distribution ratio of 70:10:20. The next step involved creating a classification model using IndoBERT, and adjusting hyperparameters to optimize its performance. Lastly, the model was evaluated using the testing data, employing various parameters such as accuracy, precision, recall, and F1-score. For a comprehensive overview of the research design, please refer to Fig. 1.

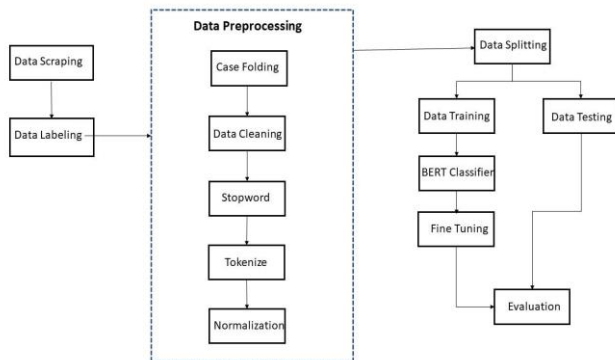


Fig. 1. Research design.

A. Data Scraping and Labeling

The dataset utilized in this research was obtained from the Google Play website. Data collection was performed using scraping techniques, employing the Python programming language and the Google-play-scraper library. The authors specifically collected review data from healthcare applications such as Alodokter, Halodoc, and Klikdokter, amassing a total of 9.310 user reviews in September 2022. To avoid the biases data, we removed any personally identifiable information (PII) from the reviews.

As the data was initially unlabeled, a manual data labeling process was conducted to facilitate the subsequent classification task; it is a very important process because the deep learning model will learn from the pattern of the given dataset.

The dataset was divided into two classes: positive and negative class. Following the completion of the labeling process, the dataset comprised 4.950 positive reviews and 4.360 negative reviews, so the dataset used is quite balanced. Detailed information regarding the datasets used in this study can be found in Table I.

TABLE I. DATASET

No	Class	Data
1	Positive	4.950
2	Negative	4.360
	Total	9.310

B. Preprocessing Data

Preprocessing is a crucial step in converting raw data into a format suitable for classification input [17]. This process involves five stages, namely case folding, data cleaning, stopword removal, tokenization, and normalization [18]. Case folding denotes a textual transformation operation that converts all capital letters within a string to a lowercase, to render the comparison and processing of text more consistent. Data cleansing constitutes a critical preprocessing step for natural language processing (NLP) pipelines. NLP involves manipulating and analyzing human language; hence the quality of the input data can substantially influence the performance and efficacy of NLP systems. This process aims to prepare the classification input by eliminating unwanted elements. Various actions are performed during data cleaning, such as removing unique characters, usernames, hashtags, punctuation, emojis, and excessive spaces, resulting in a dataset containing only words.

The next phase is known as the stopword process and entails removing common words that appear frequently but have no significant meaning. In computational linguistics and textual data analysis pipelines, function words considered uninformative are frequently used. These terms, known as stopwords, are eliminated prior to subsequent processes because they contribute marginally to the semantic content. Typically, they contain high-frequency words such as "the," "is," "and," "a," "an," "in," "of," etc. The precise stopword lexicon varies based on the objective of natural language processing and the examined language. The rationale behind the eradication of stopwords is the reduction of the dimensionality of textual data, which can accelerate processing

and improve the efficacy of specific natural language processing techniques, such as text categorization and information extraction. By removing these common terms, the emphasis transfers to more informative content words that can provide more meaningful distinctions. To accomplish this, a stop-list dictionary containing these words is compiled. Therefore, data containing stop list terms are removed to improve sentiment analysis performance.

Following this, the tokenization process is applied to convert each data entry into individual tokens, where each token typically corresponds to a single word [19], [20]. Lastly, data normalization is performed to address the presence of non-standard words commonly found in Google Play reviews. This conversion of non-standard words into standard ones is essential for improving classification performance

C. Data Splitting

After the data enters preprocessing, it will be divided into three parts: training data, validation data, and testing data [21]. Each part serves a different function: data training is used to create models, data validation reduces overfitting, and data testing evaluates the created models. The proportion of data sharing is 70% for the training set, 20% for the validation set, and 10% for the test set.

D. BERT Fine Tuning

BERT is a pre-training model that has undergone extensive training on a vast amount of data. The process of creating a BERT model involves two key steps: pre-training and fine-tuning. During pre-training, the model is trained on unlabeled data using various pre-training tasks. Subsequently, the BERT model is fine-tuned using labeled data from downstream tasks, starting with the pre-trained parameters. Despite commencing with the same pre-trained parameters, each downstream task results in a well-tuned model [12]. Fig. 2 illustrates the fine-tuning process on the pre-trained BERT model. The versatility of the BERT technique has been demonstrated through various studies aimed at addressing research gaps. For instance, researchers have successfully improved accuracy with transformer-based models when dealing with large, complex documents [22]. Additionally, BERT-based text classification has been enhanced by incorporating additional sentences and domain knowledge [23]. Notably, the impact of these improvements has been particularly evident in high-resource languages like English.

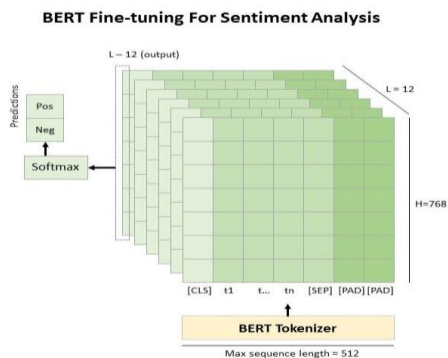


Fig. 2. BERT fine tuning.

To train pre-training models effectively, it is crucial to use specific languages. Since our dataset consists of user reviews in Indonesian, it necessitates pre-training models tailored for the Indonesian language. As a result, the BERT-based model has undergone significant enhancements, leading to the development of IndoBERT [24]. This improved version is built upon the Indonesian vocabulary, achieved by modifying the Huggingface framework. IndoBERT has been meticulously trained on an extensive dataset comprising over 220 million words, sourced from various Indonesian platforms, including Indonesian Wikipedia, news articles from Kompas, Tempo, Liputan6, and Korpus Web Indonesia. The training process involved running IndoBERT through 2.4 million steps or 180 epochs, taking approximately two months to complete. The positive attributes of IndoBERT motivated us to utilize this model for classifying Indonesian app reviews. For this study, we specifically employed "IndoBERT-base-p1," which represents one variant of the IndoBERT model [8].

E. Evaluation

Evaluation is a technique used to determine a model's classification aptitude. This study's model evaluation employs a confusion matrix that generates true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values. Multiple metrics, including accuracy, sensitivity, specificity, and precision, as well as the F1-Score, are employed to evaluate the implemented model. The accuracy formula has been shown in equation (1), the recall formula has been shown in equation (2), the precision formula has been shown in equation (3), and the F1-Score formula has been shown in equation (4).

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1 - Score = \frac{2 \times (Precision \times recall)}{recall+precision} \quad (4)$$

III. RESULT AND DISCUSSION

This research utilizes IndoBERT transfer learning, a technique that leverages a pre-trained model to address new problems of a similar nature. In this study, we employ IndoBERT as the pre-trained model, which stands for Indonesia Bidirectional Encoder Representations from Transformers, built using the PyTorch framework. IndoBERT is a transformers-based model, derived from Bert Base with 12 hidden layers, tailored specifically for monolingual Indonesian language tasks [25].

The investigation utilized a dataset comprising 9.310 samples, which were categorized into three subsets: training, validation, and test data. Fig. 3 illustrates the data labeling, with 4.950 samples carrying a positive label and 4.360 samples carrying a negative label. It is evident that positive or "good" reviews constituted 53.2% of the data, while negative reviews accounted for 46.8%.

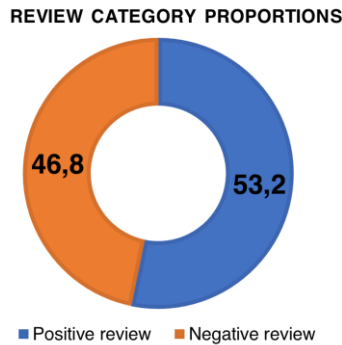


Fig. 3. Review category proportions.

The author employed 10 epochs for the training process. Observing the results, it becomes evident that utilizing 10 epochs yields commendable accuracy, depicted by a noticeable upward trend in the curve. Additionally, this study utilized a learning rate of 1e-6, as a parameter for Adam's optimizer. This parameter was chosen based on experimental induction. It is worth noting that the appropriate learning rate varies from case to case, as learning rates that are too large or too small can lead to suboptimal solutions. The learning rate typically ranges from 0 to 1, with higher values facilitating faster training but not necessarily guaranteeing more optimal results. Therefore, careful selection of the learning rate value is vital to achieve the best possible outcomes. Fig. 4 illustrates the training results against the dataset, showcasing the performance curve in relation to the chosen parameters.

According to Fig. 4, the curve exhibits a pronounced upward trend to the right, suggesting a well-trained model. Additionally, the model trained with the new dataset undergoes evaluation to determine its performance against the dataset. To assess the model's effectiveness, a confusion matrix is employed in this study. The evaluation of the model against the testing data is depicted in Fig. 5.

According to the observations from Fig. 3, the model demonstrates excellent predictive capabilities. Notably, the values for true positive and true negative are significantly higher than those for false positive and false negative. The study's results indicate an impressive accuracy score of 96%, an F1-Score of 95%, a Recall of 96%, and a Precision of 95%.

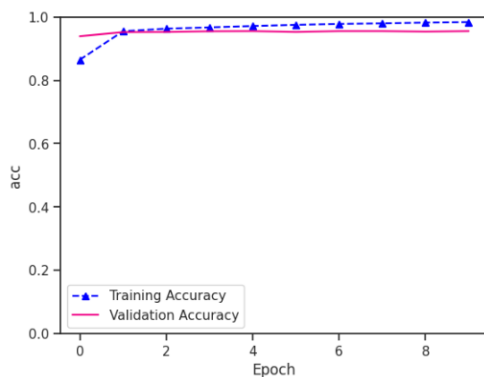


Fig. 4. Training history.

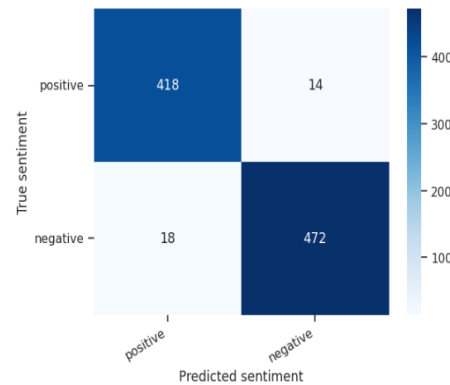


Fig. 5. Confusion matrix.

This study employs a different methodology than previous studies, such as Pandesenda et al., who conducted sentiment analysis on Alodokter data extracted from the Google Play Store in 2020 [26]. This procedure employs Fast Large-Margin, which yields an accuracy of 92.33%. Mehta et al., using Bidirectional LSTM, identify healthcare sentiment analysis from Twitter data with an accuracy of 80.88% in a separate study [27]. A comprehensive overview of these comparisons can be found in Table II.

TABLE II. COMPARISON WITH OTHER STUDIES

No	Researchers	Method	Accuracy
1	Pandesenda et al.,	Fast Large-Margin	92.33%
2	Mehta et al.,	Bidirectional LSTM	80.88%
3	Our Study	IndoBERT-base-p1	96%

The reasons for specific projections in the context of sentiment analysis are critical for various reasons. (1) Sentiment analysis provides interpretability, bridging the gap between the model's sophisticated computations and human perception of emotion. (2) Users and stakeholders have a right to know why certain decisions are being made, especially when those decisions impact their experiences or choices. (3) Domain experts can provide insights into why certain linguistic patterns may carry particular sentiment connotations in the given language or culture. The limitation of this research is that the model was trained using IndoBERT, which is specifically designed for Bahasa Indonesia content and has not been tested with other languages.

IV. CONCLUSION

In conclusion, this study focused on conducting sentiment analysis of Indonesian text using the transfer learning technique with the IndoBERT pre-trained model. The research was based on a dataset containing 9.310 reviews, each labeled as either positive or negative. During the training process, 10 epochs were used along with Adam's optimizer, employing a learning rate of 1e-6. The evaluation of the model yielded impressive results, with a high accuracy score of 96%, an F1-Score of 95%, a Recall of 96%, and a Precision of 95%. These findings underscore the effectiveness of transfer learning with IndoBERT as a robust approach for sentiment analysis of Indonesian text. If the dataset used increases, with reference to the current high accuracy value, there is a possibility that the performance will decrease but not significantly.

By contributing to the advancement of natural language processing research for the Indonesian language, this study holds significant value. The applications of this technique are diverse and can prove beneficial in areas like opinion mining, social media analysis, and market research. To further enhance the model's capabilities, future research may explore parameter optimization and evaluation with larger and more diverse datasets, thereby increasing its generalizability.

ACKNOWLEDGMENT

The author expresses gratitude to the Universitas Muhammadiyah Surakarta for providing research support, enabling the completion of this study. This research is fully funded by Riset Muhammadiyah (RisetMu).

REFERENCES

- [1] R. Alturki and V. Gay, "Usability Attributes for Mobile Applications: A Systematic Review," 2019, pp. 53–62. doi: 10.1007/978-3-319-99966-1_5.
- [2] [M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA: ACM, Aug. 2004, pp. 168–177. doi: 10.1145/1014052.1014073.
- [3] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken DW, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," in *6th International Conference on Sustainable Information Engineering and Technology 2021*, New York, NY, USA: ACM, Sep. 2021, pp. 258–264. doi: 10.1145/3479645.3479679.
- [4] M. Hassenzahl, "Experience Design: Technology for All the Right Reasons," Morgan & Claypool Publishers.
- [5] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, doi: 10.1561/15000000011.
- [6] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artif Intell Rev*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3.
- [7] S. Sah, "Machine Learning: A Review of Learning Types," pp. 1–7, 2020.
- [8] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," Sep. 2020.
- [9] F. Y. A'la, "Indonesian Sentiment Analysis towards MyPertamina Application Reviews by Utilizing Machine Learning Algorithms," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol. 5, no. 1, pp. 80–91, 2022.
- [10] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training," pp. 1–12, 2018.
- [11] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2018, pp. 328–339. doi: 10.18653/v1/P18-1031.
- [12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.
- [13] M. A. Fauzi, "Random Forest Approach for Sentiment Analysis in Indonesian Language," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 1, p. 46, Oct. 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [14] D. Alita, S. Priyanta, and N. Rokhman, "Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter," *Journal of Information Systems Engineering and Business Intelligence*, vol. 5, no. 2, p. 100, Oct. 2019, doi: 10.20473/jisebi.5.2.100-109.
- [15] R. P. Nawangsari, R. Kusumaningrum, and A. Wibowo, "Word2Vec for Indonesian Sentiment Analysis towards Hotel Reviews: An Evaluation Study," *Procedia Comput Sci*, vol. 157, pp. 360–366, 2019, doi: 10.1016/j.procs.2019.08.178.
- [16] L. H. Suadaa, I. Santoso, and A. T. B. Panjaitan, "Transfer Learning of Pre-trained Transformers for Covid-19 Hoax Detection in Indonesian Language," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, p. 317, Jul. 2021, doi: 10.22146/ijccs.66205.
- [17] M. L. L. Wijerathne, L. A. Melgar, M. Hori, T. Ichimura, and S. Tanaka, "HPC Enhanced Large Urban Area Evacuation Simulations with Vision based Autonomously Navigating Multi Agents," *Procedia Comput Sci*, vol. 18, pp. 1515–1524, 2013, doi: 10.1016/j.procs.2013.05.319.
- [18] R. Kusumaningrum, I. Z. Nisa, R. P. Nawangsari, and A. Wibowo, "Sentiment analysis of Indonesian hotel reviews: from classical machine learning to deep learning," *International Journal of Advances in Intelligent Informatics*, vol. 7, no. 3, p. 292, Nov. 2021, doi: 10.26555/ijain.v7i3.737.
- [19] N. Bahrawi, "Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based," *Journal of Information Technology and Its Utilization*, vol. 2, no. 2, p. 29, Dec. 2019, doi: 10.30818/jitu.2.2.2695.
- [20] F. Y. A'la, Hartatik, N. Firdaus, M. A. Safi'ie, and B. K. Riasti, "A Comprehensive Analysis of Twitter Data: A Case Study of Tourism in Indonesia," in *2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS)*, IEEE, Aug. 2022, pp. 85–89. doi: 10.1109/APICS56469.2022.9918757.
- [21] Merfat. M. Altawaier and S. Tiun, "Comparison of Machine Learning Approaches on Arabic Twitter Sentiment Analysis," *Int J Adv Sci Eng Inf Technol*, vol. 6, no. 6, p. 1067, Dec. 2016, doi: 10.18517/ijaseit.6.6.1456.
- [22] C. Liao, T. Maniar, S. N, and A. Sharma, "Techniques to Improve Q&A Accuracy with Transformer-based models on Large Complex Documents," pp. 1–8, 2020.
- [23] S. Yu, J. Su, and D. Luo, "Improving BERT-Based Text Classification With Auxiliary Sentence and Domain Knowledge," *IEEE Access*, vol. 7, pp. 176600–176612, 2019, doi: 10.1109/ACCESS.2019.2953990.
- [24] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.
- [25] S. L. Sariwening and Azhari, "IndoBERT: Transformer-based Model for Indonesian Language Understanding," in *Master Thesis*, Yogyakarta, 2020.
- [26] I. Pandesenda, R. R. Yana, E. A. Sukma, A. Yahya, P. Widharto, and A. N. Hidayanto, "Sentiment Analysis of Service Quality of Online Healthcare Platform Using Fast Large-Margin," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, IEEE, Nov. 2020, pp. 121–125. doi: 10.1109/ICIMCIS51567.2020.9354295.
- [27] A. Mehta, S. Virkar, J. Khatri, R. Thakur, and A. Dalvi, "Artificial Intelligence Powered Chatbot for Mental Healthcare based on Sentiment Analysis," in *2022 5th International Conference on Advances in Science and Technology (ICAST)*, IEEE, Dec. 2022, pp. 185–189. doi: 10.1109/ICAST55766.2022.10039548.

The Medical Image Denoising Method Based on the CycleGAN and the Complex Shearlet Transform

ChunXiang Liu¹, Jin Huang², Muhammad Tahir^{3,*}, Lei Wang^{4,*}, Yuwei Wang⁵, Faiz Ullah⁶

School of Resources and Environmental Engineering, Shandong University of Technology, Zibo, Shandong, China¹

School of Computer Science and Technology, Shandong University of Technology, Zibo, Shandong, China^{2,4,5}

Department of Computer Science, Mohammad Ali Jinnah University, P.E.C.H.S, Karachi, Sindh, Pakistan^{3,6}

Abstract—Medical image denoising plays an important role for the noise in the medical images can reduce the visibility, thereby affecting the diagnostic results of the doctors. Although good results have been achieved by the well-known deep learning-based denoising methods for their strong ability of learning, the loss of structural feature information and the well preservation of the edge information have not attracted considerable attention. To deal with these problems, a novel medical image denoising method based on the improved CycleGAN and the complex shearlet transform(CST) is proposed. The CST is used to construct the generator to embed more feature information in the training process and the denoising process is modeled to adversarial learn the mapping between the noise-free image domain and the noisy image domain. With the mechanism of the recurrent learning from the CycleGAN, the proposed method does not need the paired training data, which obviously speeds up the training and is more convenient than other classical methods. By comparing with five state-of-the-art denoising methods, experiments on the open dataset fully prove the accuracy and efficiency of the proposed method in terms of the visual quality and the quantitative PSNR, SSIM, and EPI.

Keywords—Medical image; image denoising; CycleGAN; complex shearlet transform

I. INTRODUCTION

Medical imaging techniques play the vital role in modern disease diagnosis, for they are the disruptive tools to observe the internal structure and functional information of human body. For example, the computed tomography(CT) can show the clear structure of the fracture [1] and the PET can effectively detect and distinguish the cancer or the normal metabolism of the lung [2]. Though great success has been achieved, the main challenge comes from the possible noise or artifacts during imaging procedure, which may result in the unexpected diagnostic errors, even the death. For example, the noise will largely affect the results of image reconstruction [3]. Thus, the effective denoising methods are highly needed to be the fundamental and mandatory step of the medical imaging pre-processing or the further applications.

Nowadays, many advanced denoising methods have been proposed, all of which can be generally divided into four categories: the filter-based methods [4], the model-based methods [5], the multi-scale geometric transform-based methods [6] and the deep learning-based methods [7, 8]. For the filter-based methods, they typically implement the low-pass filters to replace the noisy or suspected pixel by their

locally averaging value or energy in the neighboring region. The Gaussian filter, median filter and diffusion filter are the common methods in the early days. However, these methods are easy to produce the results with low contrast. Then, the bilateral filter [9], non-local filter [10], guarding filter [11], the block matching and 3D collaborative filtering (BM3D) [12] are successively proposed. They improve the denoising results, but are limited to the great diversity of the noise and the setting of the parameters, such as the height and width of the searching window. The model-based methods treat the denoising process to be a special mathematical model, for example, G. Gilboa et al. proposed to use the partial differential equations to describe the evolution of an image in time, and the solution of these equations are adapted to remove the noise and preserve the details [13]. Usually, the good results can be obtained, but the computational complexity is too high to implement in the real time application.

In recent years, for the low computational complexity and the superior properties in the frequency domain, a large number of work under the multi-scale geometric transform-based methods have been popularly reported, which decompose the noisy images into multi-resolution and different directions in each scale and then do the operations on the coefficients by the threshing scheme, considering the correlation of them, or the combination with the filters. According to the proposed time, the commonly used decomposition tools include the wavelet transform, curvelet transform, contourlet transform, non-subsampled contourlet transform, shearlet transform, non-subsampled shearlet transform [14, 15]. For example, A. Halidou et al. reported a new review on the wavelet transform based medical image denoising methods, which compare the performance of the typical wavelet, such as the discrete wavelet, Harr wavelet, and Dual-Tree complex wavelet [16]. P. S. Negi and D. Labate proposed a novel denoising method based on the discrete shearlet transform for CT images [17] and X. He et al. proposed the medical image denoising methods based on the non-subsampled version of the shearlet transform [18]. They decompose the input image into sub-images with different frequency bands and perform the denoising process for each sub-image separately, and then recompose the denoising sub-images into the results. It not only has good denoising effect and fast speed, but also has strong robustness, and can be applied in practical scene. A very good review of the multi-scale geometric transform-based methods on different image modalities can be found in [19, 20]. The benefits of the multi-

scale geometric transform-based methods are obvious since the features can be easy to capture in different scales, but they usually suffer from the disadvantages that the operations on the transformed coefficients may not match the distribution of the specific noise in different scales and directions. Though some typical models are proposed to alleviate the drawback, such as the generalized Gaussian distribution model in the wavelet and shearlet domain [21, 22], the Hidden Markov Model in the wavelet, NSCT and shearlet domain [23-25], Gaussian scale mixture model [26], the results are still not satisfying.

With the great breakthrough of the deep learning theory, it has been popularly applied in the medical imaging processing domain, such as the image U-net model for the super-resolution [27], the convolutional long short-term deep network for recognition of human action [28], and the graph resnet for motor imagery classification [29], as well as the image denoising. For example, K. Zhang et al. construct the "FFDNet" model for image denoising based on the conventional neural network [30], W. Li et al. proposed to use the fast and flexible deep convolutional neural network(FFCNN) to remove the Gaussian noise [31] and K. Zhang et al. designed beyond Gaussian denoiser by the residual learning of deep CNN [32]. R. S. Thakur et al. compared the different performance of the state-of-art image denoising methods using convolutional neural networks in [33]. Recently, the good denoising results have been obtained by the generative adversarial network (GAN) model for it models the denoising task to be the game between the noisy image and denoising image, which is implemented by the learning strategy on the generative and the discriminator network [34]. Furthermore, to suppress the influence of the diversity of the noise and control the sampling variables, the conditional generative adversarial networks (CGAN) is proposed for removing the noise of the low-dose CT images [35]. The deep learning-based methods outperform the other methods for their strong representation and generalization ability of the deep level features. Though great success has been achieved, the deep learning-based methods also suffer from some obstacles, such as the large amount of training data, the selection of the pooling functions for the specific model and the unpredictable interpretability of the deep features.

On the other hand, comprehensively considering the advantages and disadvantages of the above methods, simultaneously using the multiscale feature and the deep features may be a good way to deal with their drawbacks. Very recently, some impressive works have been reported in this domain. For example, Z. Lyu et al. constructed the "NSTBNet" model based on the non-subsampled shearlet transform and a broad convolutional neural network to remove spatially variant additive Gaussian noise [36], C. Gu et al. combine the GAN and LSTM models for 3D reconstruction of Lung Tumors from CT Scans[37], Q. Song et al. proposed the multimodal sparse transformer network (MMST) to remove the external noise in the task of the automatic speech recognition by using the mechanism of sparse self-attention [38] and B. Jiang et al. constructed the so-called "EFFNet" model for image denoising by enhancing the transformed

frequency features with dynamic hash attention [39]. Inspired by the above work, a novel image denoising method based on the complex shearlet transform and the cycle-consistent adversarial networks (CycleGAN) is developed to improve the denoising performance.

The main contribution of this research work is as follows:

Firstly, a simplified but efficient cycle-consistent adversarial network is constructed. Compared with other deep learning models, it does not need a large amount of pairwise training data with labels. So, the accuracy and robustness, stability is high.

Secondly, the image denoising is modeled to be the problem of the adversarial learn; the mapping between the noise-free image domain and the noisy image domain. As the state-of-the-art multiscale representation tool, the complex shearlet transform is employed to construct the image generator, which is able to preserve the significant and important characteristics well.

Finally, five state-of-the-art denoising methods are conducted to prove its effectiveness and accuracy. Experimental results demonstrate it produces the best denoising results both in the qualitative and quantitative analysis.

The paper is structured into several sections. Section I introduces the background of the denoising methods. Section II describes the related work on the CycleGAN and the CST. Section III presents the details of the whole proposed method. Section IV conducts the experiments and discussions. Section V finally presents the conclusion and discusses the further plan.

II. THE RELATED WORK

A. The CycleGAN

The CycleGAN model is a very typical model to deal with the problem of the image to image translation in the vision and graphics domain, whose goal is to train a useful mapping between the source domain and the target domain without the paired or aligned input-output data set.

According to [40], the principle of the CycleGAN is based on two core concepts: the basic GAN model and the Cycle Consistency Loss. The GAN is used to generate images in the target domain similar to the given training data, while the cycle consistency loss encourages the generated images to be returned to the original images in the source domain.

As shown in Fig. 1, the CycleGAN consists of two mappings functions G and F, and their associated discriminators D_x and D_y . Different from the GAN, CycleGAN contains two generators and two discriminators, where one generator converts the data from the source domain to the target domain and the other generator converts it back to the source domain. The discriminators are used to determine whether the transformed data are true or false in the two directions.

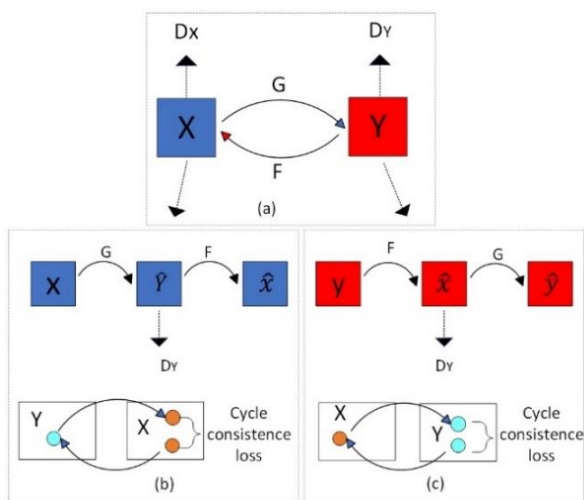


Fig. 1. The structure of the CycleGAN model.

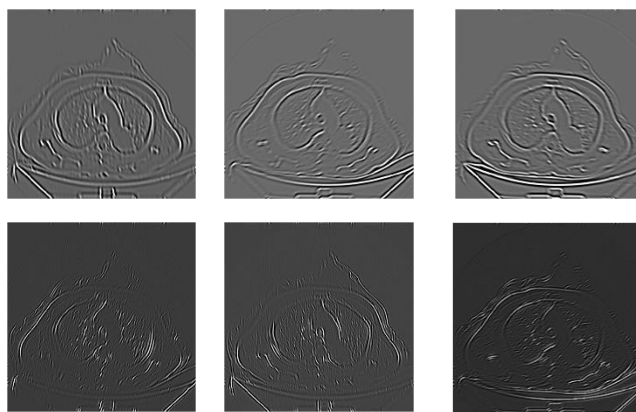
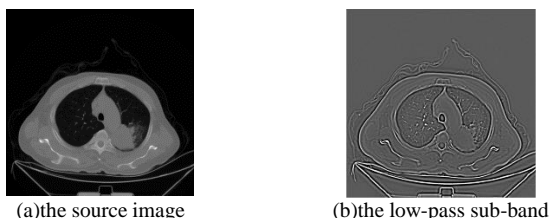
For the training process, a generator and a discriminator network are trained separately in each domain. The purpose of the discriminator is to determine whether the generated data is realistic, while the generator is to generate a more realistic data to deceive the discriminator. The input of the generator is the data from the source domain and the output is the data from the target domain. The output of the discriminator is a probability value that indicates whether the data is real or generated.

During the training process, the generator is encouraged to generate more realistic data by calculating the difference between the output of the generator and the data in the target domain. Two cycle consistency losses are used to regularize the outputting mapping.

B. The Complex Shearlet Transform

As state-of-the-art multi-scale geometric transform tool, the complex shearlet transform is especially suitable to represent the local feature of the images by using the phase and amplitude information. It has many unique characteristics. For example, the different discrete shearlet transform, the CST is proposed with strict mathematical theory guarantee to meet the Parsval frame. Furthermore, though it has the similar property of shift invariance with the non-subsampled shearlet transform, it has the simpler implementation and higher computational efficiency. In addition, the CST has stronger direction selectivity.

Actually, the discrete implementation of the CST is realized by using the Laplace pyramid for multi-resolution analysis, and the multi-scale partition filters to get the directions. How to implement the CST is not the research hotspot in this paper, more details can be found in [41]. Fig. 2 shows an example of the CST.



(c) the high-pass sub-bands at the first and second level

Fig. 2. An example of the CST.

III. THE PROPOSED METHOD

The main purpose is to make the full use of the advantage of the CycleGAN, that is, its training does not require one-to-one image samples. Only the two types of image domains are ok. Specifically speaking, for the proposed method, it does not need the sample labels to guide the training process, but only the set of images containing noise and without noise are required. It greatly enhances the generalization and makes the network more effective to avoid the overfitting phenomena in learning the mapping from the noise-containing image domain to the noise-free image domain.

A. The Whole Process

The proposed model is shown in Fig. 3.

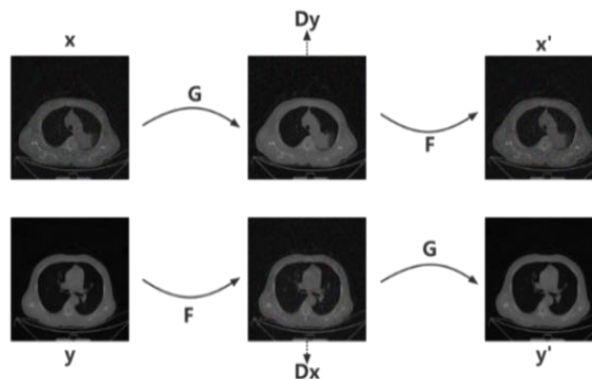


Fig. 3. The process of the proposed model.

As shown below, it mainly consists of two generators and two discriminators; X, Y is noisy images in the X and Y domain, respectively. Images in the X domain can be generated by the generator G, and then reconstructed back to X domain by generator F. Similarly, images in the Y domain can be generated by the generator F, and then reconstructed back to the Y domain by the generator G. The discriminators Dx and Dy play a discriminatory role to ensure the migration of the images.

B. The Process of Image Generation

The CycleGAN model is proposed to solve the image translation problem, so two generators are needed to realize

the transformation between two domains. However, For the image denoising problem, the aim is mainly to solve the mapping from noise-containing to noise-free without caring about the mapping from noise-free to noise-containing, so the two styles of the generators in CycleGAN are simplified in our method, and a separate noise extractor is used to realize the mutual transformation between noise-containing and noise-free images [42].

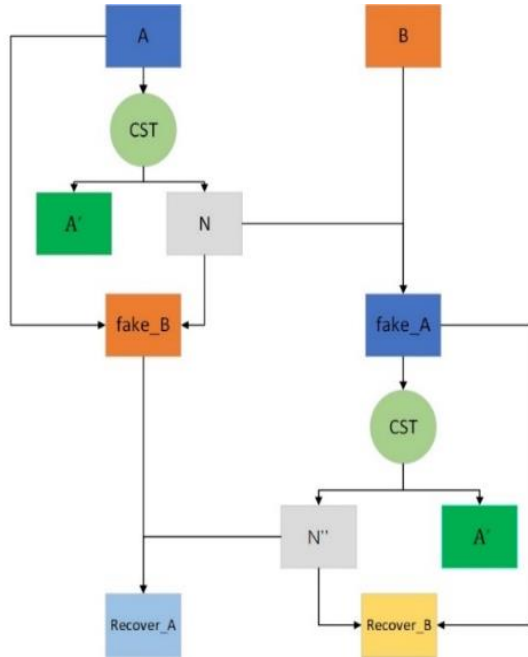


Fig. 4. The procedure of the image generation.

As shown in Fig. 4, there are six types of images involved in the generating procedure, and the CST is used to produce them to input more feature information in the training. Firstly, let A be the noisy sample image in the dataset and B be the noiseless sample; then, A is input through the CST extractor to get the noise component N, the noise image A is subtracted from the noise component N to get the denoised image fake_B, the noiseless image B is added to the noise component to get the generated noise image fake_A. The fake_A is passed through the CST noise extractor again to get the noise component N', fake_A is subtracted from the noise component N' to get the secondary noisy image. The noisy image B is added with the noise component to get the generated noisy image fake_A', the noise component N' is obtained by passing fake_A' through the CST noise extractor again, the noisy image recovered_B is obtained by subtracting fake_A' from the noise component N', the noisy image recovered_B is obtained by adding fake_B with the noise component N', and the noisy image recovered_A is obtained by adding fake_B with the noise component N'. recovered_A, and the noise-free sample image A is outputted by the CST noise extractor to predict the noise component in the noise-free image, and the ideal output value should be 0.

C. The Loss Function

In the proposed method, three models are needed to update the parameters, i.e. the noise extractor G, discriminator DA and discriminator DB. According to the basic CycleGAN, the

loss of the two discriminators consists of the discriminant error to determine whether the image is real or the generated. And the loss of the generator is composed of three losses, i.e. the loss_GAN, loss_identity and loss_cycle. In our method, the discriminant error is also maintained, and in order to improve the stability of the model and speed up the training, a new denoising loss (noted as loss_denoise) is added to the training of the model. More details on the calculation of the losses can be found in the following description.

1) Consistency loss is obtained from the final output images of A and the generator G, and the final output images of B and generator F. In an ideal state, the final output images between A and B should be identical, so the difference between them is used to be the consistency loss.

2) The adversarial loss is the opposite to the discriminatory loss of the discriminator, which represents the ability of fake_A and fake_B to deceive the discriminator. So, the correct judgment of the discriminator is used to be the adversarial loss.

3) The cyclic consistency loss is obtained from the images recovered_A and recovered_B generated by adding noise and removing noise from the generated images fake_A and fake_B again and the original images A and B. Recovered_A and recovered_B should be similar to A and B respectively to the maximum extent in order to ensure the noise is successfully removed without affecting other information. Therefore, the difference between them is used to be the cyclic consistency loss.

4) The denoising loss similar to it is in the general image denoising model. Thus, the difference between the noise-containing image after passing through the noise extractor and the noise-free image is used to be the denoising loss.

The calculation of the four types of losses can be divided into two categories, one is to calculate the error, and the other one is to calculate the difference between two images, which can be represented by the Mean Squared Error (MSE) and Mean Absolute Error (MAE) in the following equations.

$$MSE(y, y') = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (1)$$

As the calculation of the adversarial loss and the discriminator loss needs to be compared with the output of the discriminator, so y in Equation (1) is the target value with 0 or 1, y' is the output of the discriminator, and n is the number of a batch in the training, which is calculated uniformly for the output of the whole batch.

$$MAE(u, u') = \frac{\sum_{i=1}^n |u_i - u'_i|}{n} \quad (2)$$

For the other type of loss, it is necessary to compare the magnitude of the difference between the two images, so a pixel-by-pixel comparison is required. In Equation (2), u is the real image, u' is the generator-generated image, and n is the number of a batch in training.

After obtaining the four losses, the loss $loss_G$ of the noise extractor model can be obtained by Equation (3) to (7).

$$Loss_A = (loss_indentifyA + loss_indentifyB) * a \quad (3)$$

$$Loss_B = (loss_GANA + loss_GANB) * b \quad (4)$$

$$Loss_C = (loss_cycleA + loss_cycleB) * c \quad (5)$$

$$Loss_D = loss_denoise * d \quad (6)$$

$$Loss_G = Loss_A + Loss_B + Loss_C + Loss_D \quad (7)$$

In the above equation, a, b, c and d are the weighting coefficients.

After the CST, the low-frequency sub-band images usually contain few noise components, so the $loss_denoise$ accounts for a relatively small proportion of the loss, and it mainly relies on the consistency loss to ensure that the original image information is not lost. For high-frequency sub-band images, they contain more noise components. Thus, the $loss_denoise$ ratio should be adjusted upward to focus on noise removal, and the ratio of the other losses should not be adjusted downward too much, ensuring that the high-frequency details of the texture in the image not be removed by any mistake.

IV. THE EXPERIMENTS AND DISCUSSION

A. Experiment Setting

The experimental platform is the CentOS Linux with Intel Xeon Silver and the NVIDIA Tesla P100. All the codes are implemented by the PyTorch and OpenCV.

The peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and edge preservation index (EPI) are used to be the performance metrics. To save space, how to compute them can be found in [43, 44].

The experiments are designed to be two parts: training the CycleGAN denoising model and verifying the effectiveness of the denoising algorithm. According to [32, 33], the deep learning-based methods outperform the traditional methods, such as the Wiener filtering, polynomial regression, or the wavelet denoising, so the proposed method is compared with five state-of-the-art denoising methods, i.e. the NSST-BM3D model [18], the FFDNet model [30], the FFCNN model [31], the GAN model [35] and the NSTBNet model [36]. The parameters, such as the size of the convolutional layer, the network depth, are set to be the same as they are reported in the corresponding literature. For the implementation of proposed model, the basic structure of the CycleGAN is used get the best performance by tuning the parameters according to [40].

All the images can be downloaded from the public data set LIDC-IDRI [45], and 5000 images are selected in the

experiments. All of them are added the 10%, 15%, 20%, 25%, and 30% Poisson noise [46].

B. Results and Discussion

Table I shows the average PSNR values obtained by the different methods under the five noise levels of 10%, 15%, 20%, 25%, and 30%. It can be seen that the proposed method gets the best value. Compared with NSST-BM3D, FFCNN, and FFDNet, the improvements are more obvious when the noise is at the higher level. Compared with NSTBNet and GAN, the PSNR values are also higher.

TABLE I. THE PSNR VALUE OF DIFFERENT METHODS

Level	NSST-BM3D	FFCNN	FFDNet	NSTBNet	GAN	CycleGAN
10%	29.84	30.02	30.10	30.13	30.20	30.38
15%	28.01	28.18	28.29	28.36	28.39	28.63
20%	26.15	26.22	26.30	26.35	26.38	26.44
25%	25.12	25.24	25.28	25.30	25.31	25.44
30%	23.43	23.51	23.64	23.66	23.71	23.95

TABLE II. THE SSIM VALUE OF DIFFERENT METHODS

Level	NSST-BM3D	FFCNN	FFDNet	NSTBNet	GAN	CycleGAN
10%	0.872	0.875	0.875	0.882	0.875	0.891
15%	0.801	0.796	0.806	0.813	0.807	0.876
20%	0.762	0.763	0.776	0.783	0.784	0.794
25%	0.706	0.723	0.724	0.735	0.740	0.755
30%	0.663	0.703	0.710	0.715	0.719	0.723

TABLE III. THE EPI VALUE OF DIFFERENT METHODS

Level	NSST-BM3D	FFCNN	FFDNet	NSTBNet	GAN	CycleGAN
10%	0.94	0.95	0.95	0.96	0.97	0.98
15%	0.90	0.89	0.91	0.92	0.91	0.94
20%	0.83	0.846	0.86	0.87	0.87	0.89
25%	0.80	0.82	0.83	0.85	0.84	0.86
30%	0.70	0.71	0.73	0.75	0.77	0.81

Table II and Table III show the average SSIM and EPI values of different methods at the five different noise levels, respectively. The CycleGAN value is higher than that for the other methods. When the noise level increases, its advantages will slowly manifest, especially when the noise level is 20% and 25%, the SSIM is significantly higher than the other methods and when the noise level is 20%, the EPI get the best value.

The reason is that during the training process, the denoising extractor in the proposed method makes full use of the CST features via the low-pass and high-pass sub-bands coefficients. The consideration of the geometric and structural features of the source can be well maintained in the final result and guarantee the good value of the SSIM and EPI.

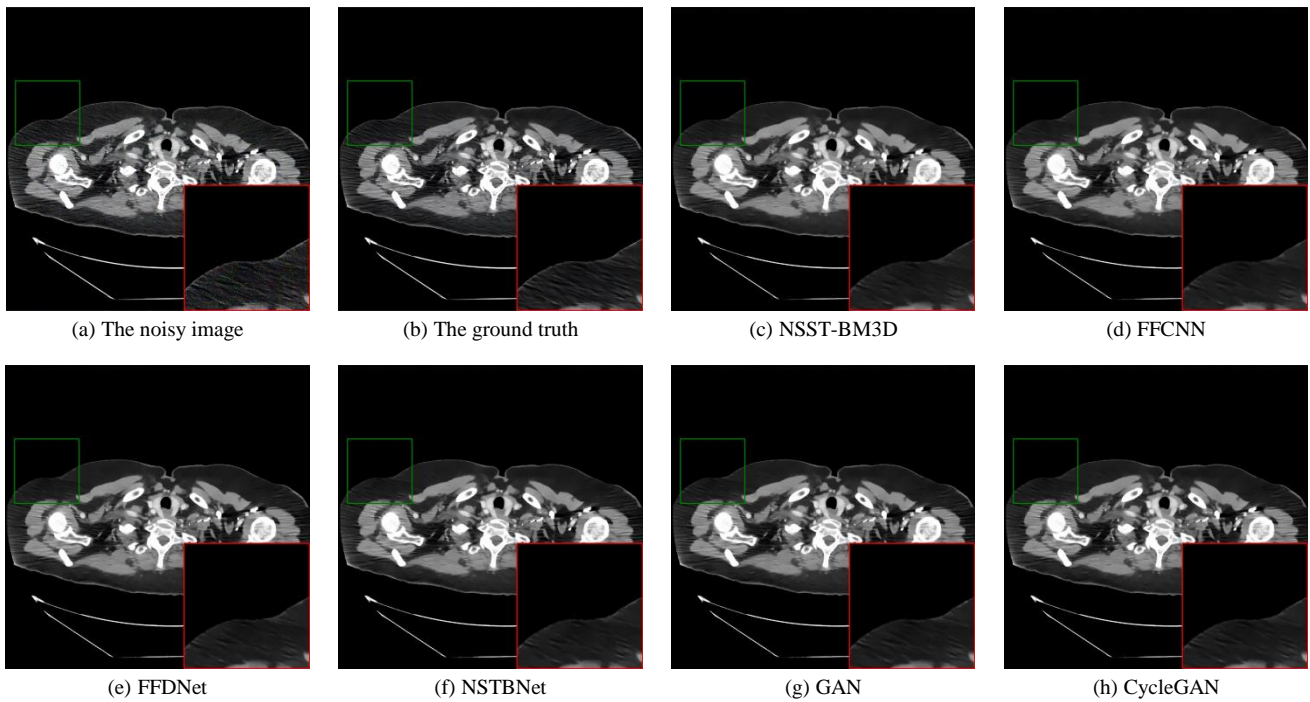


Fig. 5. The denoising results at the 10% noise level.

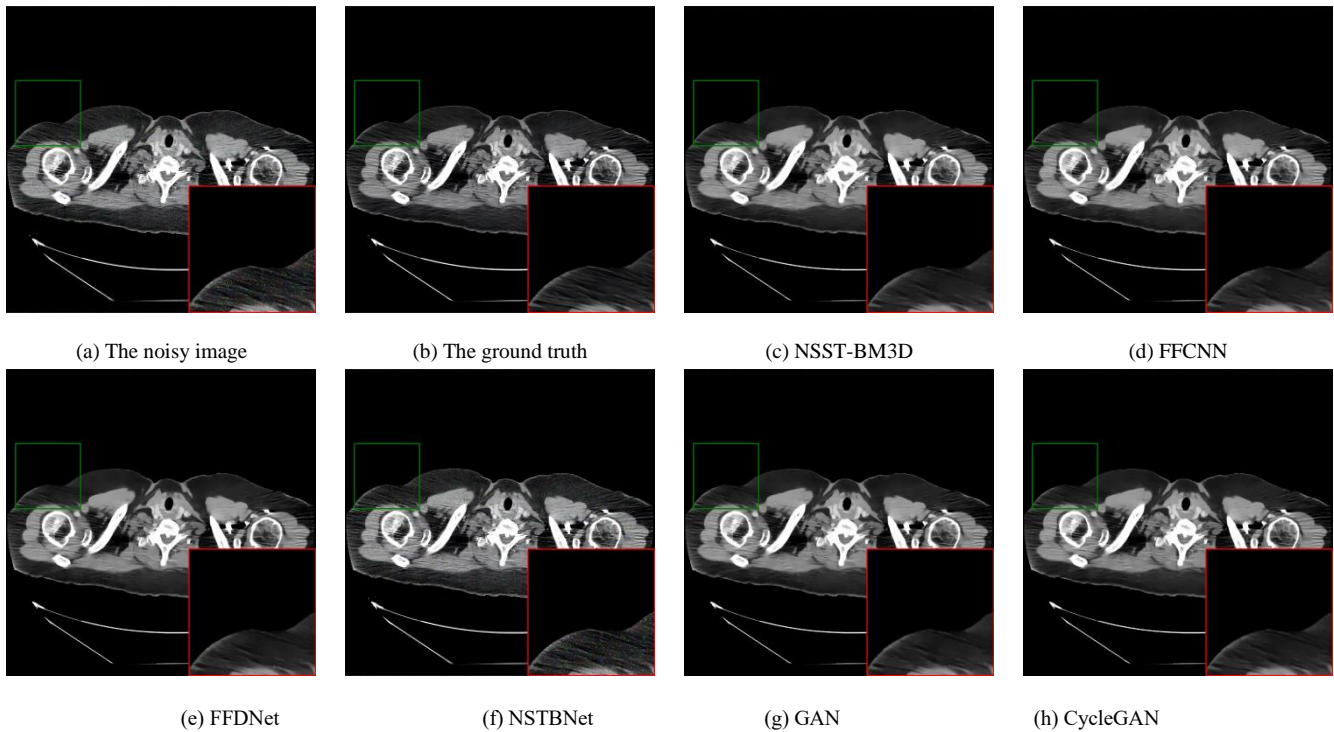


Fig. 6. The denoising results at the 15% noise level.

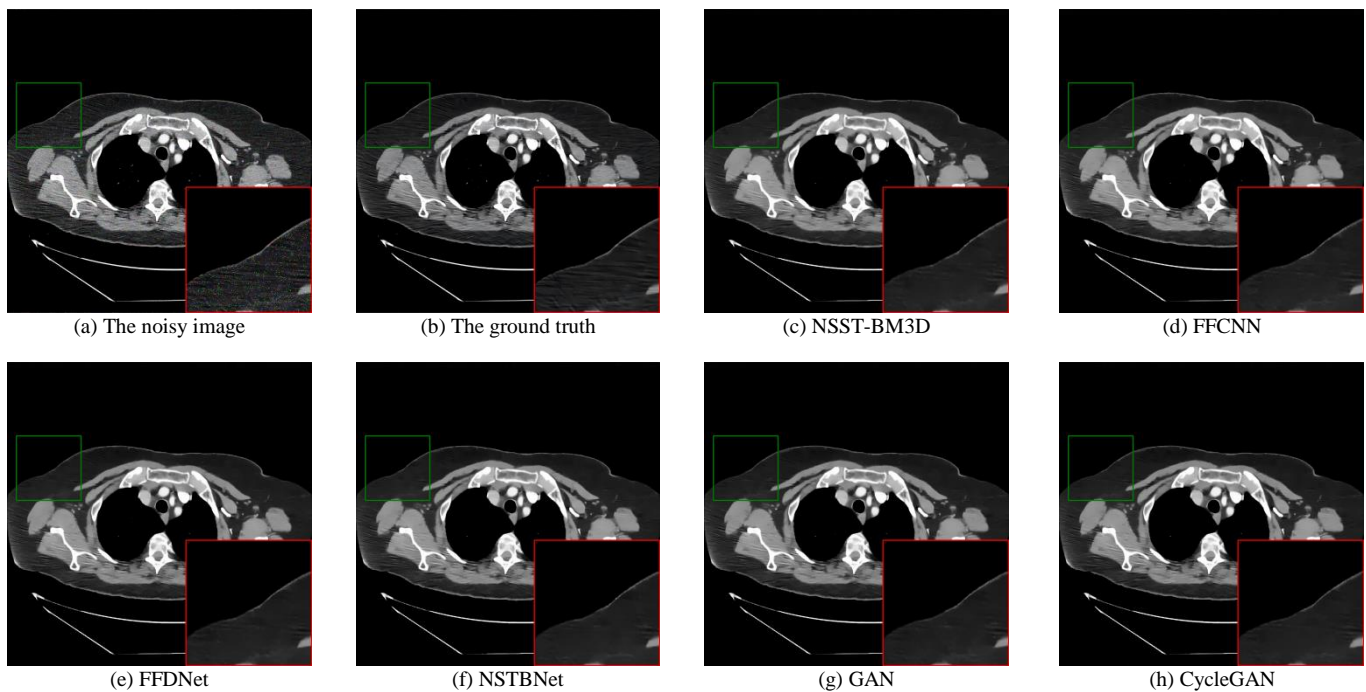


Fig. 7. The denoising results at the 20% noise level.

In Fig. 5 to Fig. 7, the visual results of the different methods are shown. Due to space limitations, only three groups of the experimental images are presented. The area marked by the red squares is enlarged at the same location. Compared with the ground truth and noisy images, the results of the CycleGAN method are more clear and they perform better in maintaining more feature details than other methods. According to the enlarged area, the details can be compared. The edge part of CycleGAN denoising image is more smooth and clearer, and the detail part is almost equivalent to it is in the ground truth.

Medical image denoising is very important in kinds of medical imaging processing tasks. After the optimization, the proposed model can be applied into the object detection, segmentation, and classification tasks enabled by the denoising techniques mentioned in this paper.

V. CONCLUSION

An effective medical image denoising method based on the improved CycleGAN model and the complex shearlet transform is proposed. The main idea is to use the multi-scale decomposition property of the CST and the principle of the recurrent learning of the GAN. The advantages mainly locate at the strong ability of the extracting the important structure and edge information of the noisy images and training an effective cycle GAN model. Compared with five state-of-the-art denoising methods on the open dataset, the validity and accuracy are fully demonstrated.

In future, we will discuss with some medical experts to implement more experiments on the data from different imaging modalities, such as the MRI, PET and Ultrasound, and consider their feedback to validate the effectiveness.

ACKNOWLEDGMENT

This study is supported by: A project ZR2021MF017 supported by Shandong Provincial Natural Science Foundation; a project 2023RKY01015 supported by the Key R&D Program of Shandong Province, China; a project ZR2020MF147 supported by Shandong Provincial Natural Science Foundation.

REFERENCES

- [1] M. Diwakar, M. Kumar, "A review on CT image noise and its denoising," *Biomedical Signal Processing and Control*, vol. 42, pp.73-88, April 2018.
- [2] J. Gong, J. Guan, C. Liu and J. Qi, "PET image denoising using a deep neural network through fine tuning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 153-161, March 2019.
- [3] J. Huang, L. Wang, M. Tahir, T. Cheng, X. Guo, Y. Wang and C. Liu, "The Effective 3D MRI reconstruction method driven by the fusion strategy in NSST domain," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol.14, no.4, pp.709-715, April 2023.
- [4] E. Yahaghi, M. Mirzapour, A. Movafeghi and B. Rokrok. "Interlaced bilateral filtering and wavelet thresholding for flaw detection in the radiography of weldment," *The European Physical Journal Plus*, vol.135, pp.42-52, January 2020.
- [5] F. Ashouri, and M. R. Eslahchi, "A new PDE learning model for image denoising," *Neural Computing and Applications*, pp. 8551-8574, March 2022.
- [6] A. Vyas, J. Paik, "Applications of multiscale transforms to image denoising: Survey," 2018 International Conference on Electronics, Information, and Communication (ICEIC), Honolulu, HI, USA, pp. 1-3, January 2018.
- [7] S. Roy, S. Imran Hossain, M. Akhand and K. Murase, "A robust system for noisy image classification combining denoising autoencoder and convolutional neural network," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 9, no.1, pp.224-235, September 2018.

- [8] S. Sagheer, and S. George, "A review on medical image denoising algorithms," *Biomedical Signal Processing and Control*, vol. 61, pp.102036, June 2020.
- [9] B. Zhang, J. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise removal," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 664-678, May 2008.
- [10] Y. Wu, S. Li. "A novel fusion paradigm for multi-channel image denoising," *Information Fusion*, vol.77, pp.62-69, August 2021.
- [11] Y. Do, Y. Cho, S H. Kang and Y. Lee, "Optimization of block-matching and 3D filtering (BM3D) algorithm in brain SPECT imaging using fan beam collimator: Phantom study," *Nuclear Engineering and Technology*, vol. 54, no. 9, pp.3403-3414, 2022.
- [12] H. Singh, S. Kommuri, A. Kumar and V. Bajaj, "A new technique for guided filter based image denoising using modified cuckoo search optimization," *Expert Systems with Applications*, vol. 176, pp.114884, August 2021.
- [13] G. Gilboa, N. Sochen and Y. Zeevi, "Image enhancement and denoising by complex diffusion processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1020-1036, August 2004.
- [14] K. Allard, G. Chen and M. Maggioni, "Multi-scale geometric methods for data sets II: geometric multi-resolution analysis," *Applied and Computational Harmonic Analysis*, vol. 32, no. 3, pp. 435-462, May 2012.
- [15] Q. Ren, B. Zhou, L. Tian and W. Guo, "Detection of COVID-19 with CT images using hybrid complex shearlet scattering networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 194-205, January 2022.
- [16] A. Halidou, Y. Mohamadou, A. Ari and E. Zacko, "Review of wavelet denoising algorithms," *Multimedia Tools and Applications*, <https://doi.org/10.1007/s11042-023-15127-0>, April 2023
- [17] P. S. Negi, and D. Labate, "3-D discrete shearlet transform and video processing," *IEEE Transactions on Image Processing*, vol. 21, no. 6, pp. 2944-2954, June 2012.
- [18] X. He, C. Wang, R. Zheng, Z. Sun and X. Li, "GPR image denoising with NSST-UNet and an improved BM3D," *Digital Signal Processing*, vol. 123, pp.103407, April 2022.
- [19] M. Hu, B. Sun, X. Kang and S. Li, "Multiscale structural feature transform for multi-modal image matching," *Information Fusion*, vol.95, pp.341-354, July 2023.
- [20] Y. Liu, S. Liu and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, no. 4, pp. 147-164, July 2015.
- [21] M. Hashemi, S. Beheshti, "Adaptive bayesian denoising for general gaussian distributed signals," *IEEE Transactions on Signal Processing*, vol. 62, no. 5, pp. 1147-1156, March 2014.
- [22] L Wang, B Li and L Tian, "EGGDD: An explicit dependency model for multi-modal medical image fusion in shift-invariant shearlet transform domain," *Information Fusion*, vol. 19, pp. 29-37, September 2014.
- [23] M. M. Ichir, A. Mohammad-Djafari, "Hidden markov models for wavelet-based blind source separation," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1887-1899, July 2006.
- [24] X. Wang, R. Song, Z. Mu and C. Song, "An image NSCT-HMT model based on copula entropy multivariate Gaussian scale mixtures," *Knowledge-Based Systems*, vol. 193, pp.105387, April 2020.
- [25] X. Wang, Y. Liu and H. Yang, "Image denoising in extended shearlet domain using hidden markov tree models," *Digital Signal Processing*, vol. 30, pp. 202-113, July 2014.
- [26] P. Gupta, A. Krishna Moorthy, R. Soundararajan and A. Bovik, "Generalized gaussian scale mixtures: A model for wavelet coefficients of natural images," *Signal Processing: Image Communication*, vol. 66, pp. 87-94, August 2018.
- [27] A. Kalluvila, "Super-Resolution of Brain MRI via U-Net architecture," *International Journal of Advanced Computer Science and Applications*, vol.14, no.5, pp.26-31, May 2023.
- [28] A. Saif, E. Wollega and S. Kalevela, "Spatio-Temporal features based human action recognition using convolutional long short-term deep neural network," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol.14, no.5, pp.1-15, May 2023.
- [29] Y. Xia, J. Dong, D. Li, K. Li, J. Nan and R. Xu, "An adaptive channel selection and graph resnet based algorithm for motor imagery classification," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol.14, no.5, pp.241-248, May 2023.
- [30] K. Zhang, W. Zuo and W. Zhang. "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol.27, no.9, pp.4608-4622, September 2018.
- [31] W. Li, H. Liu and J. Wang, "A deep learning method for denoising based on a fast and flexible convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol.60, pp.1-13, April 2021.
- [32] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang. "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol.26, no.7, pp.3142-3155, July 2017.
- [33] R. Thakur, R. Yadav and L. Gupta, "State-of-art analysis of image denoising methods using convolutional neural networks," *IET Image Processing*, vol.13, no.13, pp.2367-2380. November 2019.
- [34] A. Creswell and A. Bharath, "denoising adversarial autoencoders," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 4, pp. 968-984, April 2019.
- [35] M. B. de Almeida, L. F. Alves Pereira, T. I. Ren, G. D. C. Cavalcanti and J. Sijbers, "The gated recurrent conditional generative adversarial network (GRC-GAN): application to denoising of low-dose CT images," 34th SIBGRAP Conference on Graphics, Patterns and Images (SIBGRAP), Gramado, Rio Grande do Sul, Brazil, pp. 129-135, October 2021.
- [36] Z. Lyu, Y. Chen, Y. Hou and C. Zhang, "NSTBNet: toward a non-subsampled shearlet transform for broad convolutional neural network image denoising," *Digital Signal Processing*, vol.123, pp.103407, April 2022.
- [37] C. Gu, H. Gao, "Combining GAN and LSTM models for 3D reconstruction of lung tumors from CT scans," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol.14, no.5, pp.378-388, May 2023.
- [38] Q. Song, B. Sun and S. Li, "Multimodal sparse transformer network for audio-visual speech recognition," *IEEE Transactions on Neural Networks and Learning Systems*, pp.1-11, April 2022.
- [39] B. Jiang, J. Li, H. Li, R. Li, D. Zhang and G. Lu, "Enhanced frequency fusion network with dynamic hash attention for image denoising," *Information Fusion*, vol. 92, pp.420-434, April 2023.
- [40] J. Zhu, T. Park, P. Isola and A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp.2242-2251, October 2017.
- [41] M. Wang, C. Sun and A. Sowmya, "Complex shearlets and rotary phase congruence tensor for corner detection," *Pattern Recognition*, vol.128, pp.108606, August 2022.
- [42] I. Anokhin, K. Demochkin, T. Khakhulin, G. Sterkin, V. Lempitsky and D. Korzhenkov, "Image generators with conditionally-independent pixel synthesis," *The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.14278-14287, June 2021.
- [43] E. Reehorst and P. Schniter, "Regularization by denoising: clarifications and new interpretations," *IEEE Transactions on Computational Imaging*, vol.5, no.1, pp.52-67, March 2019.
- [44] K. Egiazarian, M. Ponomarenko, V. Lukin and O. Leremiev, "Statistical evaluation of visual quality metrics for image denoising," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.6752-6756, November 2018.
- [45] S. Kollem, K. Reddy and D. Rao, "Improved partial differential equation-based total variation approach to non-subsampled contourlet transform for medical image denoising," *Multimedia Tools and Applications*, pp.2663-2689, January 2021.
- [46] R. Harper, S. Flammia and J. Wallman, "Efficient learning of quantum noise," *Nature Physics*, vol.16, pp.1184-1188, August 2020.

Comparing Scrum Maturity of Digital and Business Process Reengineering Groups: A Case Study at an Indonesia's State-Owned Bank

Gloria Saripah Patara, Teguh Raharjo

Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia

Abstract—Bank XYZ, an Indonesia's state-owned bank, has been conducting business and digital transformation throughout its organization. Based on a recent McKinsey survey, less than 30% of organizations succeed in transformation. Fast changing business requirements and various technology-based initiatives enforce the organization to employ an Agile methodology and Scrum, to cope with the situation. Group Grp-DGT and Grp-BPR are two groups in Bank XYZ that manage their projects using Scrum. Grp-DGT develops digital projects, whereas Grp-BPR develops Business Process Reengineering (BPR) projects. Scrum maturity in both groups needs to be appraised to promote sustainability in the long run. Comparing Scrum maturity between digital and BPR projects has not been done in the previous works, especially in a state-owned bank in Indonesia. This research will help the organization through the research output which are Scrum maturity level at both groups and proposed recommendations to improve Scrum practices. The other organizations can benefit from the recommendations as well. Scrum maturity model (SMM) is used to appraise the practices, while Agile Maturity (AMM) is used to calculate the maturity rating. From this research, it is found that Grp-DGT has reached maturity level 5 (optimizing), whereas Grp-BPR is still at level 1 (initial). Based on assessment results and Scrum guides, the recommendations are then drafted. There are 15 recommendations proposed to Grp-BPR to reach level 2 and onwards.

Keywords—Transformation; scrum; digital project; BPR project; scrum maturity model; agile maturity model

I. INTRODUCTION

Business transformation has become a catch-all term for years now. It refers to how organizations reach their fullest potential. It aims to improve overall performance by generating more revenue, reducing operational expenses, and improving both customer satisfaction and productivity among employees [1]. Business processes are reengineered to be more efficient and optimized in terms of the way customer performs their financial transactions. In line with this business transformation, digital transformation has been progressing in organizations. It utilizes cutting-edge technologies to boost the current operations and to create new business opportunities [2–5]. Based on McKinsey survey [6], there are more than 80% organizations that have undertaken efforts to apply digital transformation in the past five years, and less than 30% succeed it. To win these transformations, the organizations have to manage their projects effectively. Agile project management fits the condition. It has been used in business

process improvement [7, 8]. It is rapid and adaptive to change, builds effective communication among all stakeholders, brings customer into the team, and promotes a self-managed team. It also delivers software rapidly and incrementally to compete with fast-changing market [9–11].

Bank XYZ has been aggressively performing both transformations through its two groups (or divisions in other organizations). Those groups are Digital Group (Grp-DGT) and Business Process Reengineering Group (Grp-BPR). Grp-DGT is a group developing digital projects, whereas Grp-BPR is a group developing BPR (Business Process Reengineering) projects. Bank XYZ needs Scrum maturity assessment as a part of evaluation of the current software development process in both groups. This gap raises two questions: What is the current maturity level of Bank XYZ? How does Bank XYZ improve its level? To answer these questions, this research intends to compare Scrum maturity level in Grp-DGT and Grp-BPR. It also recommends improvement in Scrum practices based on the assessment results. These recommendations can be used to support product delivery sustainability in the long run.

There are few previous case studies in Indonesia that conduct Scrum maturity assessment. Scrum maturity model (SMM) is used to perform an assessment to Scrum practices in a telecommunication company [12], an education technology startup [13], and two software development companies [14, 15]. They proposed recommendations to the organizations based on assessment results. Panjaitan et al. [14] discussed the results and the recommendations in a thorough approach. In addition, Scrum maturity level can also be compared between two groups as conducted in research [16] and [17]. Setiawan et al. [16] compared Scrum practices in a Corporate Strategy group and an Information Technology (IT) group at a telecommunication company, whereas Zelfia et al. [17] compared an IT group and a temporary unit at a state-owned bank. Comparing groups in a state-owned bank that develop digital and BPR projects has not been done previously.

Problem identification and root cause analysis are performed through direct observation and semi-structured interview. Literature study is then performed to find previous related case studies to be used as theoretical foundation and research instrument's drafting guidelines. This research will combine Scrum maturity model (SMM) and Agile maturity model (AMM) to appraise the Scrum practices and to calculate key process area (KPA) rating respectively.

Elicitation is performed by administering an online questionnaire to the respondents. Then, recommendations are proposed based on assessment results. Conclusion, limitations, and future work are also described.

This paper is constructed as follows. Section II depicts an overview of Agile methodology, Scrum framework, and Scrum maturity model. Section III explains about the research methodology utilized in this study. Section IV describes the results and the discussion related to this study's purpose including proposed recommendations for the organization. Section V shows the conclusion of this study's result, limitations, and suggestion for the future study.

II. LITERATURE REVIEW

A. Agile Methodology

Agile is a way of thinking based on values, governed by principles, and manifested in numerous practices. Based on the circumstances, agile practitioners favor certain practices over others. Agile software development was formalized in 2001 through the Agile Manifesto [9, 18].

There are four values in the Manifesto, and they are promoted in software development process [9, 14, 18–22]. Those values are (1) individuals interacting to arrive at solutions, (2) focus on delivering well-functioning software, (3) customer and developers collaborating constantly, and (4) emphasizing on responding to requirement change.

Twelve principles were derived from the values, to clarify them [9, 18, 20–22]. Those principles are (1) prioritizing customer satisfaction, (2) receiving requirements change, (3) delivering well-functioning software constantly, (4) daily interaction and collaboration between business people and developers, (5) motivating individuals to build the project, (6) using face-to-face conversations to share information to and within development team, (7) project progress is evaluated through a well-functioning software, (8) development sustainability is achieved when the sponsors, developers, and users maintaining their pace constantly, (9) constant focus to technical excellence and good design, (10) simplifying things to maximize outcome and impact, (11) self-organizing teams promotes the best designs, specifications, and architectures, and (12) the team gives thought on how to be more effective, then calibrates and consequently adjusts its behavior.

Fig. 1 illustrates the interconnection among values, principles, and practices of the Agile Manifesto.

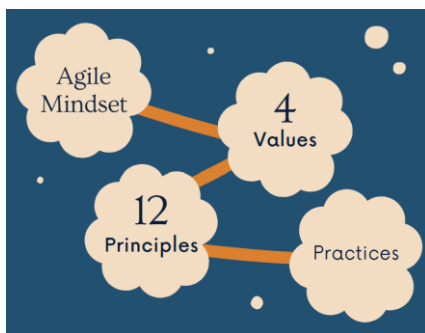


Fig. 1. The interconnection among the values, principles, and practices of the agile manifesto.

Despite the term “agile” becoming popular after the Manifesto, the viewpoints and methods have been practiced for many years before that [9, 11]. It is a superset term covering various techniques and frameworks. Fig. 2 shows the relationship among Agile and the other related terms. It is depicted as a superset term pointing to all kinds of approaches which meet the values and principles of the Agile Manifesto. Agile and the Kanban Method are shown as subsets of lean because they practice the same concepts, such as attention to value, incremental delivery, and effective process [9]. Based on a recent survey mentioned in [13, 17], Scrum is the most popular Agile approach among other approaches.

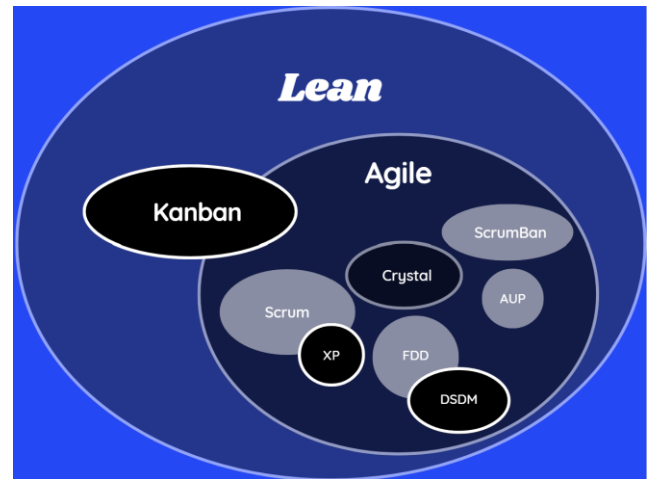


Fig. 2. Agile and other approaches [9].

B. Scrum Framework

Scrum is a simple and nimble framework that aids people, teams, and organizations in achieving goals and creating value by employing flexible approaches to solve complex problems. Empiricism and lean thinking are the foundation of Scrum. Empiricism means that the team constantly learns and improves from their past wrongdoings. Decisions and changes are made based on what the customer really needs, rather than what the developers think the customer needs. Whereas lean thinking focuses on providing benefit to the customer and assumes that anything else is inessential.

There are three primary ideas or pillars of empirical process: transparency, inspection, and adaptation [23, 24]. According to Schwaber et al. [24], transparency is mirrored through Scrum's artifacts that are visible to those performing the task. On the other hand, inspection and adaptation are implemented through four formal events in an iteration.

In Scrum, the product is delivered using an iterative, incremental approach to manage risks and to optimize predictability [24]. Commitment, courage, focus, openness, and respect are Scrum values. People's proficiency over these values determines the success of Scrum utilization throughout its process. As mentioned in [23], the process is categorized into five phases: (1) initiation, (2) planning and estimation, (3) implementation, (4) review and retrospective, and (5) release.

Based on Schwaber et al. [24], there are three roles in a Scrum development team who collectively focus their effort on a common goal, that is a product goal. The roles are

Product Owner (PO), Scrum Master (SM), and developers (DEV). VMEDU [23] categorized these three roles as core roles, and added business stakeholders, supporting services, vendors, and Scrum guidance body as non-core roles. An ideal team usually comprises of 5-9 members, 1 PO, 1 SM, and 3-7 DEVs. The team is self-organized, and each member has their own responsibilities.

The PO is the voice of business stakeholders and accountable for ensuring that the value is delivered through product increments. He or she articulates prioritized business requirements which are managed in the product backlog. The product backlog, including its items, must be visible, transparent, and understandable to the developers. The developers have specific skills to build the product. They are accountable for drafting a plan and backlog for Sprint. They also ensure deliverables quality through a Definition of Done and adapt their plan daily to meet the Sprint Goal. The Scrum Master is an individual who enables the team and the entire organization to understand what Scrum is, both theory and practice. He or she is also accountable for ensuring a proper work environment by removing impediments, so the developers can focus on delivering a high-value increment [23, 24].

Sprint is the centre of Scrum, where the team turned the business requirements into value. Fig. 3 illustrates Scrum flow for a Sprint. It is timeboxed for one to four weeks. When a Sprint concludes, it is immediately followed by a new Sprint. There are four events contained in a Sprint: Sprint planning, Daily Scrums, Sprint reviews, and Sprint retrospectives. In Sprint planning, the team discusses why this Sprint is valuable, what can be delivered, and how the selected work can be delivered. The Sprint backlog is defined in this event. Sprint goal, the selected product backlog items, and the delivery plan are part of Sprint backlog. Sprint goal is inspected daily through a Daily Scrum. The developers can synchronize their tasks, discuss potential problems, and plan for the next tasks. Definition of Done (DoD) is adhered during development. The developers then will demo the increment to stakeholders in a Sprint review event. The purpose is to obtain a review on the increment and discuss what to do next according to the current environment. The Sprint is concluded in the Sprint retrospective where future improvements are discussed [24].

Product backlog, Sprint backlog, and increment are the three artifacts mentioned in [24]. Information transparency to all team members is promoted through these artifacts, so they can be inspected, and an adaptation can be performed accordingly. A product backlog consists of ordered business requirements which are called product backlog items (PBIs). This artifact is changed based on the review or discussions with the stakeholders. The collection of PBIs which are selected to be delivered in a Sprint is called a Sprint backlog. The developers update and add more information into it along the Sprint. It must be completed to meet the Sprint goals. Increment is a delivered value which consists of the selected PBIs that have been completed in a specific Sprint. It is a steppingstone to the product goal.

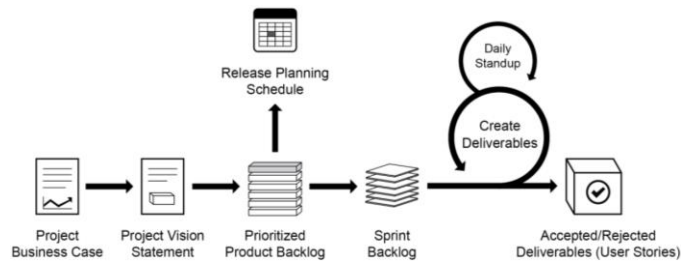


Fig. 3. Scrum flow for a sprint [23].

C. Scrum Maturity Model

According to Hutabarat et al. [25], a maturity model in a project management is a continuous process to recognize, evaluate, apply, and reassess the opportunities to improve constantly in project implementation. It is one of the organization success factors which has many types of projects, programs, and portfolios. In line with that, [14, 15] added that maturity model is a technique to evaluate the maturity level and development process capability. It continuously directs and enhances the organization's development process to avoid project failures.

The SMM refers to two maturity models, which are the AMM and the capability maturity model integration (CMMI) [12, 14]. The AMM links the Agile software development practices to maturity levels to make it simple, comprehensible, and applicable. It is designed based on Agile software development values, practices, and principles [26]. Fig. 4 depicts the AMM from an initial level to sustained level. At the initial level, an organization has not defined Agile development process clearly. At the explored level, the organization has shown more structured and complete software development practices than the first level. When an organization has practices related to customer relationship management, pair programming, communication, testing, and software quality, then it has reached the defined level. The improved level can be reached when an organization has collected of development process detailed measurement and has practiced software quality measurements. Finally, at the sustained level, an organization constantly enhances their processes through surveys and do not hesitate to have innovative initiatives [14, 26].

The CMMI is a process model that explicitly states what an organization should do to define, comprehend, and encourage behaviors that guide to improved accomplishment [27]. CMMI-DEV V1.3 mentioned that CMMI has five maturity levels: (1) Level 1 - Initial, (2) Level 2 - Managed, (3) Level 3 - Defined, (4) Level 4 - Quantitatively managed, and (5) Level 5 - Optimizing. The processes are usually ad hoc and disordered at level 1. A stable environment is usually not provided to support processes. Level 2 can be achieved when the processes are managed and performed according to documented plan. An organization achieves level 3 when processes are well described and comprehended, and are well explained in standards, procedures, tools, and techniques. At level 4, quantitative objectives for quality and process performance are established by the organization, and then utilizes them as barometer for projects management. The objectives are drafted and proposed based on the requirements

elicited from the business stakeholders. Finally, at level 5, an organization pays attention to constantly enhancing process performance through incremental and innovative processes, and technological refinement [28].

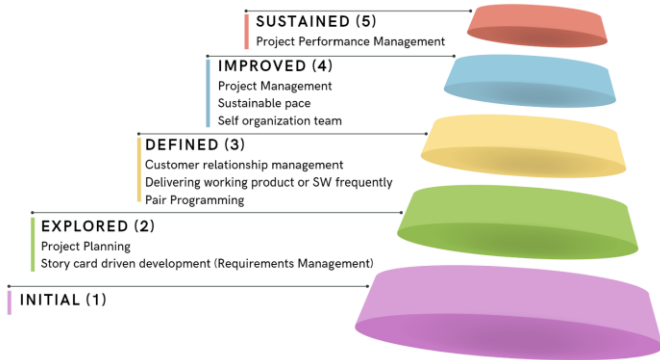


Fig. 4. Agile maturity model staged representation.

The SMM uses the same five levels as in CMMI. Its primary purpose is to guide organizations on promoting self-improvement and client’s active involvement. In addition, it also helps organizations to adopt Scrum on a staged approach by providing list of goals, objectives, practices, and metrics for every level [29]. Table I describes goals and their objectives at every level, starting from level 1 (Initial) to level 5 (Optimizing).

TABLE I. GOALS AND OBJECTIVES OF SCRUM MATURITY MODEL

Level	Code	Goals and Objectives
1 – Initial	L1	-
2 – Managed	L2.1	Basic Scrum Management
	L2.1.1	• Scrum Roles Exist
	L2.1.2	• Scrum Artifacts Exist
	L2.1.3	• Scrum Meetings Occur and are Participated
	L2.1.4	• Scrum Process Flow is Respected
	L2.2	Software Requirements Engineering
	L2.2.1	• Clear Definition of Product Owner
L2.2.2	• Product Backlog Management	
L2.2.3	• Successful Sprint Planning Meetings	
3 – Defined	L3.1	Customer Relationship Management
	L3.1.1	• Definition of Done exists
	L3.1.2	• Product Owner available
	L3.1.3	• Successful Sprint Review Meetings
	L3.2	Iteration Management
	L3.2.1	• Sprint Backlog Management
	L3.2.2	• Planned iterations
	L3.2.3	• Successful Daily Scrum
L3.2.4	• Measured Velocity	
4 – Quantitatively managed	L4.1	Unified Project Management
	L4.1.1	• Unified Project Management
	L4.2	Measurement and Analysis Management
L4.2.1	• Measurement and Analysis Management	
5 – Optimizing	L5.1	Performance Management
	L5.1.1	• Successful Sprint Retrospective
L5.1.2	• Positive Indicators	

III. RESEARCH METHODOLOGY

A. Research Stages

The objective of this research is to assess the level of Scrum maturity practices and propose recommendations for Bank XYZ’s software development process.

This research is designed to use an explanatory sequential mixed-method approach. As illustrated in Fig. 5, its stages start from problem identification to drafting suggestions for future work. The research problem is identified through an observation and semi-structured interview with a Scrum Master from group Grp-DGT and a Scrum Master from group Grp-BPR. Scrum maturity assessment has never been done in both groups, and these Scrum Masters also agreed that the assessment needs to be done to evaluate the current process. Literature study is then performed to obtain previous case studies with the same research questions. At this stage, theoretical foundations are acquired. The next stage is to construct the assessment questionnaire which is used as the research instrument. It is constructed based on SMM assessment questions. After the elicitation process, the data is analyzed using KPA rating formula from AMM. Maturity level at each goal is interpreted using this rating. The assessment result is used to find Scrum practices that need to be improved, and to draft proposed recommendations based on those findings. As the final stage, the author concludes the research and gives suggestions for future work.

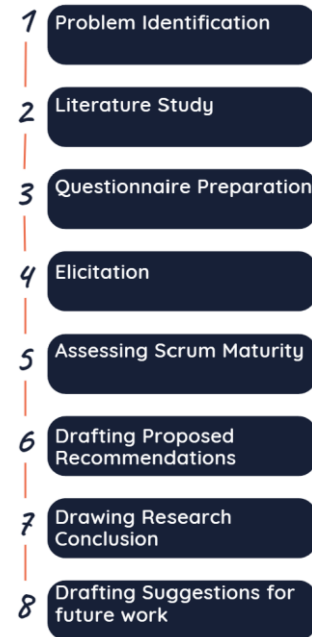


Fig. 5. Research stages of scrum maturity assessment at bank XYZ.

B. Instrument

This research uses a questionnaire as an instrument to collect data from respondents. The questionnaire is drafted based on SMM assessment questions explained in Yin et al. [29]. As described in Table II, there are 91 Scrum practices in total that will be assessed. All practices are transposed into questions which can be responded to as ‘Yes’, ‘Partially’, ‘No’, and ‘N/A’ (not applicable).

TABLE II. DETAIL COUNT OF ASSESSED SCRUM PRACTICES ON THE QUESTIONNAIRE

Goal/ Objective Code	Count	Goal/ Objective Code	Count	Goal/ Objective Code	Count
L2.1	28	L3.1	9	L4.1	1
L2.1.1	3	L3.1.1	3	L4.1.1	1
L2.1.2	9	L3.1.2	2	L4.2	2
L2.1.3	10	L3.1.3	4	L4.2.1	2
L2.1.4	6	L3.2	22	L5.1	11
L2.2	18	L3.2.1	8	L5.1.1	5
L2.2.1	5	L3.2.2	6	L5.1.2	6
L2.2.2	7	L3.2.3	5		
L2.2.3	6	L3.2.4	3		
Total of Assessed Scrum Practices	28 + 18 + 9 + 22 + 1 + 2 + 11				91

C. Elicitation

The questionnaire is drafted and distributed using Google Form. The questionnaire respondents are selected using purposive sampling technique. It is used to obtain data from the ones who understand the research problem [30]. There are four respondents who will fill out the questionnaire. They are two Scrum Masters who involved in digital projects and two Scrum Masters who involved in BPR projects.

D. Maturity Assessment

The appraisal of Scrum practices is performed using a questionnaire to obtain how practices are implemented in the projects. Data collected from the questionnaire will be analyzed using the KPA rating which is used in AMM [26]. The term “process area” in AMM is on a par to term “goal” in SMM. KPA rating can be calculated using equation as shown in (1).

$$R = \frac{\sum(Y_n) + \frac{1}{2} \sum(P_n)}{\sum(T_n) - \sum(NA_n)} \times 100\% \tag{1}$$

Where:

R = KPA rating

Y_n = ‘Yes’ responses

P_n = ‘Partially’ responses

T_n = Total assessed Scrum practices

NA_n = ‘N/A’ responses

Calculated KPA rating can be interpreted based on following categories [26]:

- 1) *Fully achieved*: 86% to 100% practices in the assessed KPA have been applied and proofs can be provided.
- 2) *Largely achieved*: 51% to 85% practices in the assessed KPA have been applied and proofs can be provided.
- 3) *Partially achieved*: 16% to 50% practices in the assessed KPA have been applied and some proofs can be provided.
- 4) *Not achieved*: 0% to 15% practices in the assessed KPA have been applied and a little or no proof can be provided.

IV. RESULT AND DISCUSSION

Scrum maturity assessment results are discussed for each level, starting from level 2 to level 5. KPA rating interpretation that will be mentioned along the discussion is coded as F (Fully Achieved), L (Largely Achieved), P (Partially Achieved), and N (Not Achieved). Scrum Masters who filled out the assessment questionnaire are also coded as X1 and X2 for the ones seated in group Grp-DGT. While Scrum Masters seated in group Grp-BPR are coded as Y1 and Y2.

A. Groups Assessment Result – Scrum Maturity Level 2

Basic Scrum Management (BSM) and Software Requirements Engineering (SRE) are two goals in Scrum maturity level 2. Table III shows the maturity level assessment results on first mentioned goal. KPA rating obtained for Grp-DGT is 95.53%. Scrum practices in four objectives listed in the table below are applied to more than 86.00% or applied almost entirely in project development. So, it can be said that BSM goal reaches Fully Achieved. On the other hand, KPA rating for Grp-BPR is 80.97%. Scrum practices in the listed objectives are applied to more than half of them, but it’s still lower than 86.00%. It means that the rating is interpreted as Largely Achieved.

TABLE III. ASSESSMENT RESULT OF BASIC SCRUM MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Scrum Roles Exist	100.00	100.00	50.00	100.00
Scrum Artifacts Exist	100.00	100.00	77.78	100.00
Scrum Meetings Occur and are Participated	90.00	100.00	60.00	81.25
Scrum Process Flow is Respected	83.33	91.67	83.33	91.67
Rating per Scrum Master	92.86	98.21	69.64	92.31
Rating per group	95.53		80.97	
Interpretation	Fully Achieved		Largely Achieved	

As shown in the Table IV, both Grp-DGT and Grp-BPR scored the same result for Software Requirements Engineering (SRE) goal assessment. The KPA rating is 94.44% which means that Scrum practices in three objectives listed in the table below are applied more than 86.00% or applied almost entirely in project development. So, it can be said that the rating for both groups in SRE goal can be interpreted as Fully Achieved.

TABLE IV. ASSESSMENT RESULT OF SOFTWARE REQUIREMENTS ENGINEERING

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Clear Definition of Product Owner	100.00	90.00	100.00	100.00
Product Backlog Management	85.71	100.00	85.71	92.86
Successful Sprint Planning Meetings	91.67	100.00	100.00	91.67
Rating per Scrum Master	91.67	97.22	94.44	94.44
Rating per group	94.44		94.44	
Interpretation	Fully Achieved		Fully Achieved	

B. Groups Assessment Result – Scrum Maturity Level 3

Customer Relationship Management (CRM) and Iteration Management (IMG) are the goals in Scrum maturity level 3. Table V shows the maturity level assessment results on first mentioned goal. KPA rating obtained for Grp-DGT is 91.66%. Scrum practices in three objectives listed in the table below are applied to more than 86.00% or applied almost entirely in project development. So, it can be said that CRM goal reaches Fully Achieved. Whereas KPA rating for Grp-BPR is 94.44% which means that Scrum practices in the listed objectives below are also applied to more than 86.00% or almost entirely in project development. So, the rating can be interpreted as Fully Achieved.

TABLE V. ASSESSMENT RESULT OF CUSTOMER RELATIONSHIP MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Definition of Done exists	100.00	100.00	100.00	100.00
Product Owner available	100.00	100.00	100.00	100.00
Successful Sprint Review Meetings	62.50	100.00	87.50	87.50
Rating per Scrum Master	83.33	100.00	94.44	94.44
Rating per group	91.66		94.44	
Interpretation	Fully Achieved		Fully Achieved	

As shown in the Table VI, assessment result for goal IMG scored slightly different at 87.74% for Grp-DGT and 87.91% for Grp-BPR. It means that Scrum practices in three objectives listed in the table below are applied more than 86.00% or applied almost entirely in project development. These ratings can be classified as Fully Achieved.

TABLE VI. ASSESSMENT RESULT OF ITERATION MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Sprint Backlog Management	68.75	75.00	75.00	81.25
Planned iterations	100.00	91.67	83.33	90.00
Successful Daily Scrum	100.00	100.00	100.00	100.00
Measured Velocity	100.00	100.00	100.00	100.00
Rating per Scrum Master	86.84	88.64	86.36	89.47
Rating per group	87.74		87.91	
Interpretation	Fully Achieved		Fully Achieved	

C. Groups Assessment Result – Scrum Maturity Level 4

Two goals in Scrum maturity level 4 are Unified Project Management (UPM), and Measurement & Analysis Management (MAM). Table VII shows the maturity level assessment results on first mentioned goal. KPA rating obtained for Grp-DGT is 100.00% which is undoubtedly interpreted as Fully Achieved. It means that Scrum practices in an objective listed in the table below are applied entirely in project development. On the contrary, KPA rating for Grp-BPR only reaches 75.00%. Scrum practices in the listed objective are applied to more than half of them, but it's still lower than 86.00%. It means that UPM goal reaches Largely Achieved. Grp-BPR's rating is a bit contrast compared to Grp-DGT's perfect rating.

TABLE VII. ASSESSMENT RESULT OF UNIFIED PROJECT MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Unified Project Management	100.00	100.00	100.00	50.00
Rating per Scrum Master	100.00	100.00	100.00	50.00
Rating per group	100.00		75.00	
Interpretation	Fully Achieved		Largely Achieved	

Assessment result for MAM goal is shown in Table VIII. Grp-DGT scored 100.00%, whereas Grp-BPR scored 87.50%. It means that Scrum practices in the objective listed in the table below are applied entirely in Grp-DGT's project development. Whereas Grp-BPR applies more than 86.00% or applies almost entirely in project development. These KPA ratings can be interpreted as Fully Achieved. Despite having the same interpretation, a Scrum Master in Grp-BPR didn't perfectly satisfy with the practices.

TABLE VIII. ASSESSMENT RESULT OF MEASUREMENT AND ANALYSIS MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Measurement and Analysis Management	100.00	100.00	100.00	75.00
Rating per Scrum Master	100.00	100.00	100.00	75.00
Rating per group	100.00		87.50	
Interpretation	Fully Achieved		Fully Achieved	

D. Groups Assessment Result – Scrum Maturity Level 5

There is only one goal in Scrum maturity level 5, that is Performance Management (PMG). Table IX shows that the KPA rating obtained for Grp-DGT is 91.66% which means that Scrum practices in two objectives listed in the table below are applied to more than 86.00% or applied almost entirely in project development. So, it can be said that PMG goal reaches Fully Achieved. Whereas KPA rating for Grp-BPR is 75.00% which means that Scrum practices in the listed objectives are applied to more than half of them, but it's still lower than 86.00%. So, the rating can be interpreted as Largely Achieved.

TABLE IX. ASSESSMENT RESULT OF PERFORMANCE MANAGEMENT

Objectives	Grp-DGT		Grp-BPR	
	X1 (%)	X2 (%)	Y1 (%)	Y2 (%)
Successful Sprint Retrospective	90.00	100.00	100.00	80.00
Positive Indicators	75.00	100.00	75.00	50.00
Rating per Scrum Master	83.33	100.00	86.36	63.64
Rating per group	91.66		75.00	
Interpretation	Fully Achieved		Largely Achieved	

KPA rating and interpretation of each goal is shown in Table X. Grp-DGT ratings are more than 85% for all Scrum maturity goals. It means that Grp-DGT assessed as Fully Achieved overall or Scrum practices are applied almost entirely in project development. Whereas Grp-BPR has three goals with KPA rating less than 86.00%, namely goal BSM, UPM, and PMG. It means that Grp-BPR needs further improvement to reach Fully Achieved overall.

TABLE X. ASSESSMENT RESULT SUMMARY

Level	Goals and Objectives	Grp-DGT	Grp-BPR
		Rating (Int.)	Rating (Int.)
2	Basic Scrum Management	95.53 (F)	80.97 (L)
	• Scrum Roles Exist	100.00 (F)	75.00 (L)
	• Scrum Artifacts Exist	100.00 (F)	88.89 (F)
	• Scrum Meetings Occur and are Participated	95.00 (F)	70.62 (L)
	• Scrum Process Flow is Respected	87.50 (F)	87.50 (F)
	Software Requirements Engineering	94.44 (F)	94.44 (F)
	• Clear Definition of Product Owner	95.00 (F)	100.00 (F)
	• Product Backlog Management	92.85 (F)	89.28 (F)
	• Successful Sprint Planning Meetings	95.83 (F)	95.83 (F)
3	Customer Relationship Management	91.66 (F)	94.44 (F)
	• Definition of Done exists	100.00 (F)	100.00 (F)
	• Product Owner available	100.00 (F)	100.00 (F)
	• Successful Sprint Review Meetings	81.25 (F)	87.50 (F)
	Iteration Management	87.74 (F)	87.91 (F)
	• Sprint Backlog Management	71.87 (L)	78.12 (L)
	• Planned iterations	95.83 (F)	86.66 (F)
	• Successful Daily Scrum	100.00 (F)	100.00 (F)
	• Measured Velocity	100.00 (F)	100.00 (F)
4	Unified Project Management	100.00 (F)	75.00 (L)
	• Unified Project Management	100.00 (F)	75.00 (L)
	Measurement and Analysis Management	100.00 (F)	87.50 (F)
	• Measurement and Analysis Management	100.00 (F)	87.50 (F)
5	Performance Management	91.66 (F)	75.00 (L)
	• Successful Sprint Retrospective	95.00 (F)	90.00 (F)
	• Positive Indicators	87.50 (F)	62.50 (L)

TABLE XI. CURRENT VS EXPECTED MATURITY

Code	Objectives	Current (%)	Expected (%)
L2.1.1	Scrum Roles Exist	75.00	86.00
L2.1.2	Scrum Artifacts Exist	88.89	86.00
L2.1.3	Scrum Meetings Occur and are Participated	70.62	86.00
L2.1.4	Scrum Process Flow is Respected	87.50	86.00
L4.1.1	Unified Project Management	75.00	86.00
L5.1.1	Successful Sprint Retrospective	90.00	86.00
L5.1.2	Positive Indicators	62.50	86.00

As illustrated at Fig. 6, there are four objectives that will be discussed further as their ratings are below the expected rating which is equal to or more than 86.00%. Those objectives are (1) Scrum roles exist, (2) Scrum meetings occur and are participated, (3) Unified project management, and (4) Positive indicators. Whereas there are three objectives that exceed the expected rating: (1) Scrum artifacts exist, (2) Scrum process flow is respected, and (3) Successful Sprint retrospective.

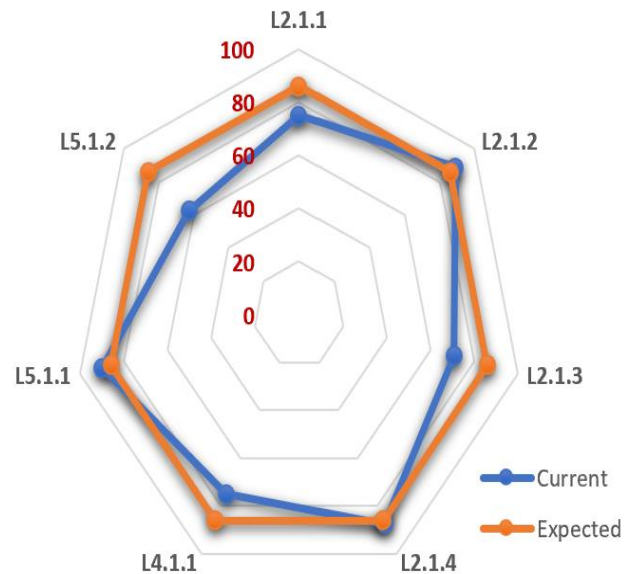


Fig. 6. Comparison of the current and expected maturity.

E. Scrum Practice Recommendation

Improvement recommendations are proposed to Grp-BPR because the assessment found that there are three SMM goals that don't reach Fully Achieved. The goals that need improvement regarding the application of the Scrum practices are Basic Scrum Management (BSM), Unified Project Management (UPM), and Performance Management (PMG). These goals are spread at SMM level 2, level 4, and level 5. Table XI describes seven objectives of the mentioned goals. Three objectives from BSM goal, one objective from the UPM goal, and two objectives from PMG goal. The rating of each objective is compared against the expected rating.

Table XII describes recommendation of improvement that can be done by Grp-BPR to reach Fully Achieved overall. The recommendations are based on SMM questionnaire responses, Scrum Guide [24], SBOK Guide [23], and A Guide to Scrum and CMMI [27].

TABLE XII. RECOMMENDATION OF IMPROVEMENT

Goals	Objectives	Recommendation
Basic Scrum Management	Scrum Roles Exist	Product Owner (PO) is a person who collect requirements from business stakeholders. He or she must ensure the team comprehends the value they are going to deliver. PO must be appointed as the project starts. He or she must know the responsibilities of a PO.
		Developers are people who build and deliver value collectively. They work based on business requirements written in a product backlog. The role must exist and staffed as the project starts.
	Scrum Meetings Occur and are Participated	Release planning is held optionally to obtain commitment over an increment delivery plan. Responsibilities, needed resources, and activities are discussed at this event.
		In release planning event, Scrum Master (SM) and PO must be present.
		Daily Scrum is one of the events that occurred in a Sprint. Developers discuss their task progress and its impediments.
		This event must be held daily on workdays to inspect progress toward Sprint goal and adapt the Sprint backlog as needed.
		Developers must attend Daily Scrum event, whereas PO and SM can attend it as developers, if they are working on Sprint backlog.
		In Sprint review event, the team presents the increment that they have built to the stakeholders. This event must be held once every Sprint.
		Sprint retrospective is held to discuss on how to better the next Sprint, in terms of the value's quality and process effectiveness. This event must be held once every Sprint.
		SM and developers must attend Sprint retrospective, whereas PO isn't mandatory to attend it.
Unified Project Management	Unified Project Management	All projects must adhere to the entire goals, objectives, and practices described in Scrum Maturity Level 2 and 3. Project Management Office (PMO) can enforce procedures to support the adherence.
Performance Management	Positive Indicators	SM coaches the team to successfully perform their tasks. SM should also remove the impediments throughout the project. Servant leadership style helps boosting the team's growth and energy level in their Scrum journey.
		A survey, at least once every Sprint, can be conducted to obtain the team's satisfaction level. This survey can also be part of Sprint retrospective.
		A survey can be conducted to obtain the stakeholders' satisfaction level. This survey can also be part of Sprint review.
		SM must ensure Scrum process has been adhered. Proper planning and task estimation should promote a work-life balance. Extra working hours must be minimized.
		SM must promote a psychologically save environment for the team. Constructive criticism and discussions in every event are welcome.

V. CONCLUSION

This research aimed to compare the Scrum maturity between two groups at Bank XYZ, namely Digital Group (Grp-DGT) and Business Process Reengineering Group (Grp-BPR). Based on assessment result, the recommendations for improvement are proposed to better Scrum practices at both groups. The following conclusions are drawn according to the research:

1) Group Grp-DGT has reached Scrum maturity level 5 (optimizing). KPA ratings of SMM goals are interpreted as Fully Achieved overall. Whereas Grp-BPR is still at level 1 (initial). Goal "Basic Scrum management" appraised as Largely Achieved with rating 80.97%. There are two objectives in this goal that don't meet the minimum rating to be appraised as Fully Achieved. Moreover, goal "Unified project management" and "Performance management" are also appraised as Largely Achieved with the same rating 75.00%.

2) Proposed recommendations for Scrum practices improvement are aimed for Grp-BPR due to its three goals are appraised as Largely Achieved. Deeper into the objective's rating, there are only four out of seven objectives whose ratings are below 86%. The recommendations are then drafted according to SMM questionnaire responses ("partially" and "no") in these four objectives. There are 15 recommendations in total for three goals consisting of nine recommendations for "Basic Scrum management", one recommendation for "Unified project management", and the last five recommendations for "Performance management". These recommendations can be used to improve Scrum practices in Bank XYZ, specifically at Grp-BPR group.

This research output can be used by Bank XYZ as a part of evaluation of the current project development process in Grp-DGT and Grp-BPR groups. Bank XYZ is also able to use it to tackle current problems in the other projects or initiatives that are not covered in this research or proposed by other groups. It would also provide aid in increasing Scrum maturity level of the next projects or Sprints in Grp-DGT and Grp-BPR groups. The other organizations, as required, would also benefit from the research output by applying the recommendations to those specific objectives.

VI. LIMITATIONS AND FUTURE WORK

This research utilized Scrum maturity model (SMM) to perform assessment to project development process at Grp-DGT and Grp-BPR groups of Bank XYZ. The elicitation was performed using a purposive sampling technique where there were four respondents participating, two respondents seated as Scrum Master at Grp-DGT, and the other two respondents seated as Scrum Master at Grp-BPR. The distributed questionnaire has 91 questions in total. There were 15 recommendations to improve four objectives that were found below the expected KPA rating.

There are some limitations of this research: (1) using only questionnaire to collect data, (2) proposed recommendations of improvement are based on SMM questionnaire responses

and Scrum guides, (3) risk impact of the unrealized Scrum practices is not investigated or being the part of the research questions, and (4) scope of study is limited to Bank XYZ.

Based on those limitations, it is suggested for the future researchers to do these works: (1) combining questionnaire, interview, and other data collection techniques to enhance the quality of the assessment results, (2) drafting the recommendation based on the combination of Scrum guides and Scrum expert judgment, (3) investigating risk impact of the unrealized Scrum practices, and (4) extending the case study scope by including some selected financial institutions.

REFERENCES

- [1] McKinsey, "What is business transformation?," Apr. 17, 2023. <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-business-transformation#/> (accessed Jun. 12, 2023).
- [2] S. Kraus, P. Jones, N. Kailer, A. Weinmann, N. Chaparro-Banegas, and N. Roig-Tierno, "Digital Transformation: An Overview of the Current State of the Art of Research," *Sage Open*, vol. 11, no. 3, p. 21582440211047576, Jul. 2021, doi: 10.1177/21582440211047576.
- [3] C. Gong and V. Ribiere, "Developing a unified definition of digital transformation," *Technovation*, vol. 102, p. 102217, 2021, doi: <https://doi.org/10.1016/j.technovation.2020.102217>.
- [4] S. Kraus, S. Durst, J. J. Ferreira, P. Veiga, N. Kailer, and A. Weinmann, "Digital transformation in business and management research: An overview of the current status quo," *Int J Inf Manage*, vol. 63, p. 102466, 2022, doi: <https://doi.org/10.1016/j.ijinfomgt.2021.102466>.
- [5] G. Vial, "Understanding digital transformation: A review and a research agenda," *The Journal of Strategic Information Systems*, vol. 28, no. 2, pp. 118–144, 2019, doi: <https://doi.org/10.1016/j.jsis.2019.01.003>.
- [6] McKinsey, "Unlocking success in digital transformations," Oct. 29, 2018. <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/unlocking-success-in-digital-transformations#/> (accessed Jun. 12, 2023).
- [7] A. Schmitt and S. Hörner, "Systematic literature review – improving business processes by implementing agile," *Business Process Management Journal*, vol. 27, no. 3, Emerald Group Holdings Ltd., pp. 868–882, 2020, doi: 10.1108/BPMJ-10-2019-0422.
- [8] P. Badakhshan, K. Conboy, T. Grisold, and J. vom Brocke, "Agile business process management: A systematic literature review and an integrated framework," *Business Process Management Journal*, vol. 26, no. 6, Emerald Group Holdings Ltd., pp. 1505–1523, Nov. 16, 2020, doi: 10.1108/BPMJ-12-2018-0347.
- [9] PMI and Agile Alliance, *Agile Practice Guide*. Pennsylvania: Project Management Institute, Inc., 2017. Accessed: Jun. 12, 2023. [Online]. Available: <https://www.pmi.org/pmbok-guide-standards/practice-guides/agile>
- [10] P. Marnada, T. Raharjo, B. Hardian, and A. Prasetyo, "Agile project management challenge in handling scope and change: A systematic literature review," *Procedia Comput Sci*, vol. 197, pp. 290–300, 2022, doi: <https://doi.org/10.1016/j.procs.2021.12.143>.
- [11] T. Raharjo and B. Purwandari, "Agile Project Management Challenges and Mapping Solutions: A Systematic Literature Review," in *Proceedings of the 3rd International Conference on Software Engineering and Information Management*, in ICSIM '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 123–129, doi: 10.1145/3378936.3378949.
- [12] N. F. Arifin, B. Purwandari, and F. Setiadi, "Evaluation and Recommendation for Scrum Implementation Improvement with Hybrid Scrum Maturity Model: A Case Study of A New Telco Product," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, 2020, pp. 178–183, doi: 10.1109/ICIMCIS51567.2020.9354311.
- [13] K. C. Abimaulana, E. K. Budiardjo, K. Mahatma, and A. Hidayati, "Evaluation of Scrum-Based Software Development Process Maturity using the SMM and AMM: A Case of Education Technology Startup," in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2021, pp. 1–5, doi: 10.1109/ICACSIS53237.2021.9631308.
- [14] I. Panjaitan and N. Legowo, "Measuring Maturity Level of Scrum Practices in Software Development Using Scrum Maturity Model," *Journal of System and Management Sciences*, vol. 12, no. 6, pp. 561–582, 2022, doi: 10.33168/JSMS.2022.0633.
- [15] N. Panjaitan and B. Hardian, "Maturity Level Analysis in Software Development Using Scrum Methodology: Xyz Startup Case Study," *Asian Journal of Social and Humanities*, vol. 1, no. 10, pp. 713–720, 2023.
- [16] J. Setiawan, F. Gunawan, T. Raharjo, and B. Hardian, "Application of Scrum Maturity Model: A Case Study in a Telecommunication Company," *J Phys Conf Ser*, vol. 1566, no. 1, p. 012050, 2020, doi: 10.1088/1742-6596/1566/1/012050.
- [17] H. Zelfia, T. Simanungkalit, and T. Raharjo, "Comparison of Scrum Maturity Between Internal and External Software Development: A Case Study at One of the State-Owned Banks in Indonesia," in *2022 1st International Conference on Information System & Information Technology (ICISIT)*, 2022, pp. 312–317, doi: 10.1109/ICISIT54091.2022.9872843.
- [18] The Agile Alliance, "Manifesto for agile software development," 2001. <http://agilemanifesto.org> (accessed Aug. 13, 2023).
- [19] A. C. Pacagnella Junior and V. R. Da Silva, "20 Years of the Agile Manifesto: on Agile Project Management," *Management and Production Engineering Review*, vol. 14, no. 2, pp. 37–48, Jul. 2023, doi: 10.24425/mper.2023.146021.
- [20] A. Kakar, "A Rhetorical Analysis of the Agile Manifesto on its 20th Anniversary," *Journal of the Southern Association for Information Systems*, vol. 10, no. 1, pp. 20–29, Feb. 2023, doi: 10.17705/3JSIS.00030.
- [21] D. Trivedi, "Agile methodologies," *International Journal of Computer Science & Communication*, vol. 12, no. 2, pp. 91–100, 2021.
- [22] S. Al-Saqqa, S. Sawalha, and H. Abdelnabi, "Agile software development: Methodologies and trends," *International Journal of Interactive Mobile Technologies*, vol. 14, no. 11, pp. 246–270, 2020, doi: 10.3991/ijim.v14i11.13269.
- [23] VMEdu, *A Guide to the SCRUM BODY OF KNOWLEDGE (SBOK® Guide)*, Fourth Edition. 2022.
- [24] K. Schwaber and J. Sutherland, "The Scrum Guide," Nov. 2020. <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf> (accessed Jun. 16, 2023).
- [25] N. Hutabarat, T. Raharjo, B. Hardian, A. Suhanto, and A. Wahbi, "PMMM Kerzner Questionnaire Validation for Project Management Maturity Level Assessment: One of the Largest Indonesia's State-Owned Banks," in *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2021, pp. 1–6, doi: 10.1109/ICACSIS53237.2021.9631325.
- [26] P. Chetankumar and M. Ramachandran, "Agile Maturity Model (AMM): A Software Process Improvement framework for Agile Software Development Practices," *International Journal of Software Engineering*, vol. 2, Jan. 2009.
- [27] CMMI Institute, *A Guide to Scrum and CMMI®: Improving Agile Performance with CMMI*. 2016.
- [28] Software Engineering Institute, "CMMI® for Development, Version 1.3," Nov. 2010.
- [29] A. Yin, S. Figueiredo, and M. Mira da Silva, "Scrum Maturity Model," Jan. 2011.
- [30] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*, Eighth Edition. Pearson, 2019.

Adaptive Learner-CBT with Secured Fault-Tolerant and Resumption Capability for Nigerian Universities

Bridget Ogheneovo Malasowe¹, Maureen Ifeanyi Akazue², Ejaita Abugor Okpako³,
Fidelis Obukohwo Aghware⁴, Arnold Adimabua Ojugo⁵, Deborah Voke Ojie⁶

Department of Computer Science, Faculty of Computing, University of Delta, Agbor. Delta State, Nigeria¹

Department of Computer Science, Faculty of Science, Delta State University Abraka, Delta State, Nigeria²

Department of Information and Communication Technology, Faculty of Computing, University of Delta, Agbor. Delta State, Nigeria³

Department of Computer Science, Faculty of Computing, University of Delta, Agbor. Delta State, Nigeria⁴

Department of Computer Science, College of Science, Federal University of Petroleum Resources Effurun, (FUPRE), Delta State, Nigeria⁵

Department of Software Engineering, Faculty of Computing, University of Delta, Agbor, Delta State, Nigeria⁶

Abstract—The post covid-19 studies have reported significant negative impact witnessed on global education and learning with the closure of schools' physical infrastructure from 2020 to 2022. Its effects today continues to ripple across the learning processes even with advances in e-learning or media literacy. The adoption and integration therein of e-learning on the Nigerian frontier is yet to be fully harnessed. From traditional to blended learning, and to virtual learning – Nigeria must rise, and develop new strategies to address issues with her educational theories as well as to bridge the gap and negative impact of the post covid-19 pandemic. This study implements a virtual learning framework that adequately fuses the alternative delivery asynchronous-learning with traditional synchronous learning for adoption in the Nigerian Educational System. Result showcases improved cognition in learners, engaged qualitative learning, and a learning scenario that ensures a power shift in the educational structure that will further equip learners to become knowledge producer, help teachers to emancipate students academically, in a framework that measures quality of engaged student's learning.

Keywords—Adaptive blended learning; computer-based test; fault tolerant design; resumption capabilities; Nigeria; FUPRE

I. INTRODUCTION

Learning is simply the process that leads to the alteration in the capability of a system causing a change in behaviour as it tries to accomplish tasks [1]. Learning thus, must yield a change or alteration in a system [2]–[4]; and thus, must equip and re-position the learner with a better knowledge to deal with task(s) at hand [5]. These changes may yield acquired skillsets, modified values, improved preferences and attitude, and new knowledge cum understanding for a learner [6]–[8] – all of which essentially improves a learner's experience and performance [9]–[11], and grants him/her the capability of new data readily, made available for future use in varying/similar instances of tasks [12]. Learning is a lifelong skillset acquired in a bid to accomplish a task over and over again [13], [14]. Learning modifies the behaviour of a learner system as it acquires the requisite skills and knowledge, which is therein stored and retrieved on demand access for use to

resolve challenges as well as assess risks, explore cum exploit challenges and opportunities [15], [16].

The covid-19 pandemic era witnessed a global lockdown of many public infrastructures with the following events namely: (a) closure of schools [17], (b) adoption of social distancing as a means to curb and reduce the spread [18], [19], (c) the migration and mobility pattern of residents from one place to another [20], and (d) adoption of nose-masks in public places [21], [22]. The post-covid-19 studies report that: (a) conventional schools were shut down to curb its widespread propagation, (b) enforcement of school shutdown resulted in various impact on the learning process for over 1.6 billion students, (c) short-term disruptions in the learning timeframe with significant negative impacts, and (d) these impacts rippled across the learning-verse, a variety of long-lasting effects in the formation of learner [23], [24].

The adoption of ICT has further bolstered and revolutionized the learning process with e-learning variants [25] as a means to help learners creatively contribute to knowledge creation (a hard feat to obtain in traditional schools with classroom settings). e-Learning today involves use of ICT-based services and electronic media-formats that can be replicated within medium that cuts across a variety of platforms in the learning process [26]. While, e-learning offers many benefits, the negative impacts of covid-19 era on education includes: (a) adoption of social distancing to curb the spread of the pandemic [27], (b) lessened migration of residents from one place to another, (c) adopting nose-masks in public places/gathering, and (d) closure of schools. Studies have reported that the covid-19-era – had significant negative-impact on the learning prospects of over 1.6 billion learners globally; and where such effects are not properly handled, may have long-lasting effects on the formation of such students.

The implications is: (a) non-access to physical infrastructure for learning, (b) yielded increased learning inequality and losses resulting from the stratified Internet access [28], [29], (c) learning disparities resulting from digital revolution/integration across a variety of learning platforms in

Nigeria (uLesson etc.), (d) the learner psychomotor and cognition health stability to adjust to blended, ICT-rich-learning paradigm as a new reality, and (e) learner adaptabilities to new complex logistics in these new paradigm, and other costs [30]–[32]. These, are determined through formative and summative testing.

II. METHODS AND MATERIALS

A. Computer-based Testing

Taking into account the changes already in place, there is a need to ensure that today's students are competent in science and technology. (Ojugo et al. [33] and Durojaye et al. [34]). The goals of science education must thus be: (a) to ensure basic science-tech literacy to enhance daily living of residents in our society today, (b) prepare learners for further training in science-tech, (c) develop skills and attitudes to prepare us for technological growth, (d) to improve learner creativity and innovation, (e) to enhance user-friendly interaction of various careers and provide knowledge applications. Fagbola et al. [35] Poor performance in science-tech can be attributed to: (1) the nature of the subject, (2) curriculum design, (3) teacher/learner characteristics, (4) the teaching style, and (5) disjointed teaching to meet examination deadlines.

To enable teachers and students harness unrestricted online access to resources in order to contribute and share knowledge with the rest of society, it is necessary to change these attitudes regarding education alongside summative testing based on CBT [36]. Technology integration with CBT is not the only thing that's going to get you into a knowledge era; rather, it leans more on the attitude of both the teacher/learner, and the requisite changes made via a paradigm shift by the teaching method as enabled by ICT. It is necessary for teachers to take on the basic issues, as well as questions about this and other topics such as: (a) teacher's literacy and awareness in mixed learning; (b) how teachers navigate a new technology together with their expectations around CBT [37].

B. Testing and Attitudinal Types

In order to enhance students' mastery of essential content, the pedagogical test is a way in which teachers use as part of their preparation for teaching feedback on continuing education and training. This also covers a wide range of formal and informal assessment methods and procedures used by teachers during the learning process in order to change the whole teaching and learning process through participatory learning, which are intended to improve learning results and retention of skills acquired throughout education [38]. Qualitative feedback for students/teachers that focuses on the content and performance is also involved, in addition to scores. Employed as assessment method, its practice presents students with clear learning targets, examples and models of strong and weak projects, regular descriptive feedbacks, the ability for self-assessment and track learning as well as set goals [39], [40].

Conversely, periodic quizzes, end-of-unit summative test, end-of-course tests and standard tests – all provide an overview of student performance at this point in time. They shall be used to assess the programme's content [41] in a

formative manner and to grade it. Albazar [42] mention the seven-principles of formative test as: (a) clarifies what good performance is about with set goals, criteria and expected standards, (b) it facilitates the development of self-assessment in learning, (c) provides quality content and engaged learnings, (d) encourages teachers and activates peer-dialogue around learning, (e) encourages positive motivational beliefs and self-esteem, (f) offers opportunities for bridging the gap between existing and desired performance, and (g) collects data for teachers that can be used as a tool for designing education and training [43], [44].

An attitude type refers to the expression of a favorable or unfavorable opinion about an individual, place, thing or event. The person's view of the target and how to talk or do things may also be referred to an attitudinal type [45], [46]. When we think about a person or something, it also means the feeling, attitude, position etc. It's a tendency or orientation in particular for the mind toward certain things [47], [48]. It refers, to the quadratic equation, to the attitude of students and teachers towards mathematics and science and technology education [49]–[51].

Iskandarov [52] Test is a *short* exam, which is grouped into formative and summative. The end of a course test is summative; while, the periodic test taken in an ongoing course is formative. To fulfil these two purposes, approximation tests are carried out to obtain a number of data on the programme's effects as well as diagnostic data for its deficiencies [53]. The Nigerian Teacher Institute groups testing into assessment methods as thus: (a) out-of-class, (b) open-book and (c) closed-book. These impact the attitude of the learner as it predisposes a learner to respond in a certain way. Younis et al. [54] attitude encompasses a range of affective behaviour with cognitive, psychomotor dimension that is measured via a self-report instrument. Attitude is a powerful motive in realization of a learner's expected goal(s).

C. Study Motivation

These challenges include (and not limited to) [55]–[57]:

1) *Paradigm shift*: The covid-19 era caught many societies, completely unprepared with the adoption/adaptation of ICT into education with Nigeria as case in point. With this shift and reform, parents and teachers had to switch their roles to become facilitators (that they were unprepared for).

2) *Questions framework*: The stratification of Internet access portends and presents network administrator(s) of the system that new questions must be uploaded onto the ensemble from time to time – to increase the pool of questions from which test-questions are drawn. Since the process is accomplished via scripting – a measure of error is unavoidably introduced with the addition of more contents onto the CBT platform.

3) *Resumption feat*: Most CBTs in administering questions – are pooled from their offline repository which are often clogged. A major issue in adopting e-learning with CBT capability is the lack of resumption ability with the learning management system (LMS) especially in lieu of physical infrastructure downtime, power outage and network failure.

4) *Poor integrity delivery outcome on the learning setting:*

The impact of the covid-19 era birthed the sudden need to adapt online teaching/learning [58], and also challenged the readiness of schools on paradigm shift to digital revolution readiness [59]. This forced adoption/adaptation of blended learning (i.e. e-learning that combines both a(synchronous) learning settings). These learning settings differed in terms of time, and the place of teaching/learning activities [60]. However, some of the challenges in the learning setting has been found to include: anonymity and social presence gaps, learning modalities, lower satisfaction, technophobia, lower cognitive achievement, disengaged participation in class of learners, fluency and slower interaction resulting from videoconferencing in e-learning [61]–[63].

D. *Proposed Adaptive Architecture for the Learner-centric Computer-based Testing (AdACoBaTe)*

Fig. 1 leverages on Dominic and Francis [26] model for e-learning as extended by [64] – which then provisions a test-based framework, which classifies learners into test groups, and proposed corresponding methods using the Sarasin model based on their learning styles namely: (a) visual learners who gain new facts and knowledge via visual inputs, (b) auditory learners who learn by listening to a teacher or teaching medium, and (c) kinesthetic/tactile learners learn via experiments and exploration (using Montessori means); however, both Hidayat and Utomo [65] commented that when a teacher is aware of the three types of learning styles outlined in Felder Silverman's 32-variant classroom activities they become more sensitive to designing exercises for improving teaching learning processes.

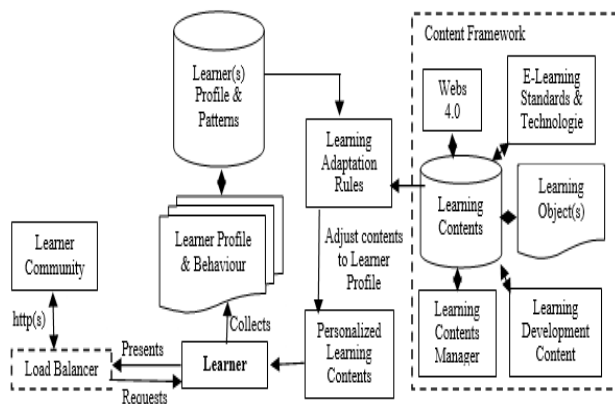


Fig. 1. AdACoBaTe: proposed adaptive content-based testing with resumption capabilities.

The workings of the ensemble is thus:

1) *The Learner Profiling Component* is that which directly interacts with the learner. It performs these tasks as thus: (a) all request and response are via HTTP/HTTPS format, (b) all request from the learning process is transmitted to a load balancing server that balances traffic on the basis of planned data structure and algorithm, (c) it collects data about learners profile and their learning behaviors, then sends them on to an adaptation model stored in a repository where they can be

formed patterns and adapted rules., and (d) it responds to learners with personalized learning contents provided by the adaptation model [66], [67].

2) *The Adaption Component* is the main processor for the experimental adaptive e-learning system. It performs these tasks: (a) it stores all learners profiles and their learning patterns into a repository, (b) it simplifies task(s) by responding to each learner, (c) it captures the browsing history, pattern and behavior of learners, and updates the captured data in its knowledge-base – helping the model to keep up with each learner and to transition between the learner profiles, (d) it uses its data content repository (or knowledge-base) about the learner’s profile to adjust the adaptation rules (and module) for each learner; And thus, identifies best learning style, learning path and learning contents suited for each identified learner, (e) it builds the personalized learning contents and hands them off to the learners’ model to present to a learner, and (f) it retrieves all contents required from the content model [68].

3) *Personalized Learning/Test-based Content* retrieves the learner’s profile and (a) provide adaptive contents from model, (b) retrieve the personalized learner’s questions in lieu of learned contents, and (c) stores them as temporary session to ease access to the contents and ensure that as the knowledge-base grows, performance is not degraded with the increase in users and the system shift between profiles [32]. Fig. 2 and 3 – show the state diagram of the ensemble as each learner log-in onto the CBT-system with resumption capabilities and fault tolerant design [69].

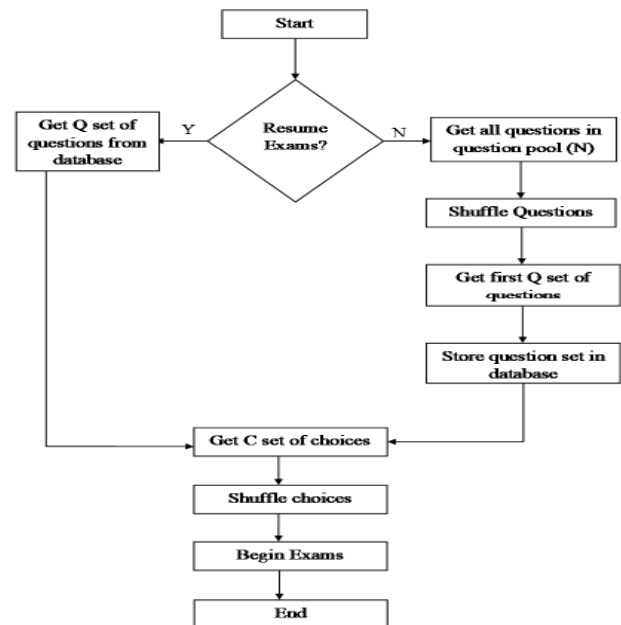


Fig. 2. Algorithm for randomized adaptive learner-based content with resumption capabilities for AdACoBaTe.

4) *The Content Component* – performs these tasks: (a) acts a repository for all the learning objects developed by a tutor cum administrator – allowing them to add/delete

questions to the pool, add/delete students list to an exam as well as monitor and manage the examination process, (b) ensures that a learner-content is based on the learning styles ratio as agreed in [70] and as suggested (i.e. visual 40%, kinesthetic 40%, and audio 20%), (c) develop contents using the latest web-tech (i.e. web 4.0 used to generate learner contents), and (d) use learning content development, management tools, and learning standards [71], [72] to proffer a system that is compatible, portable, can be ported to other systems of varying operating systems, shareable and interoperable with other e-learning systems. The preferred learning content ranges is as thus: (a) visual learning style consists of lecture materials, video lessons, animations etc, (b) kinesthetic consist of simulations, online quizzes, discussion forum, online compilations and question banks etc, and (c) audio consists of audio materials and accounts for 20-percent of learning styles.

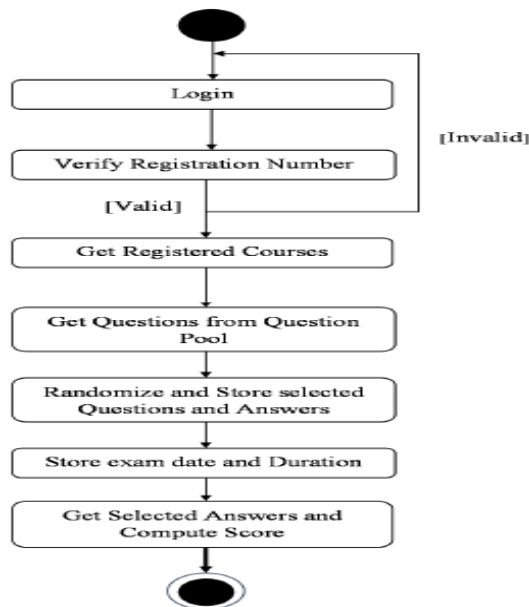


Fig. 3. Learner login with CBT access on AdaCoBaTe.

In order to prevent cheating and negligence on the part of students, a randomised question set is generally given for each student in an examination. System interfaces are easy for new computer users to use as they can navigate on a keyboard with their own set of keys. The proposed system is tolerant of failures because if there are incidents of network disruption, component failure or power outage during a test, the student's progress can be quickly restored.

a) Fault tolerance: Prior to the point of power-, component- and network-failure, system shall be capable of recover from a fault and resume the examination. Faults that occur during exams are recovered by the system. A student will not re-start the entire exam in such a case. System achieves this by tracking the entire exam, and allowing the user to resume an examination prior to the point of failure.

b) Power failure: The ensemble during examination shall be able to store each selected selection made by the student in its database and a copy will also be retained for temporary session storage. The CBT tracks client system IP-address during exams. With power fault, a user will be disconnected for an extended period pending the restore of power. If the IP address is not changed, student can re-log in and continue on the same system soon as power is restored to the machine. Also, prior to power failure, the client agent will initiate a countdown from the remainder time saved in the database. Ensemble retrieves the queries that have been entered in the database and divides them into various sets of questions, options or chosen answers.

c) Network fault: With a network fault, the client agent must not be able to send a student's newly selected answer to the orchestra if his or her computer has been disabled for an exam. To restore the network, system operation will be interrupted on the client side. The ensembles are keeping an eye on that idle time during the exam. The ensemble will give a student reallocation of idle time if it is beyond the 500-seconds threshold. Once network is back up, students are allowed to continue with their examinations from where they were before a technical failure occurred.

III. RESULTS AND DISCUSSION

A. Ensemble Performance Evaluation

Divayana [70] design seeks to ensure in e-learning systems such as public ethics, quality, accountability, nationalism etc. He categorized these feats into 10-dimensions as supported by Ojugo et al. [73]. We recast interaction with the system as a 9-dimension feat: (1) design effectiveness, (2) availability of video conference, (3) CBT readiness, (4) technical support, (5) teacher/learner use of system, (6) availability time, (7) the availability of resources, (8) the completeness of resources, and (9) data security. He used Eq. (1) to analyse the effectiveness of the learning style(s) and categorized into 4-percentiles: (a) high ranges from 81-100%, (b) sufficient from 61-80%, (c) moderate from 50-60%, and (d) poor falls below 50%.

Both categories high and sufficient – implies the learning design does not require any revision. Moderate requires some form of revision; while, the category poor implies a complete revision of the learning design. Thus, evaluation for both experts and participants yields the Table I and II respectively.

$$EP = [F/N] * 100 \quad (1)$$

Table I shows high and sufficient category ranging above 85% for all the evaluated variables by the various experts. The implication of which, is that these components do not require revisions of any kind. However, the data security component can be improved upon and agrees with [74].

TABLE I. EXPERTS' EVALUATION ON LEARNING / CBT DESIGN

Features	Experts' Evaluation Scores				
	1	2	3	4	5
Design Effectiveness	0.89	0.96	0.91	0.89	0.92
Availability of Video Conference	0.91	0.87	0.91	0.94	0.89
CBT Integration	0.75	0.89	0.79	0.89	0.91
Technical Support	0.85	0.90	0.92	0.88	0.91
Effective Usage of Framework by Teacher/Learner	0.92	0.92	0.86	0.85	0.90
Duration & Resumption Capacity	0.91	0.90	0.93	0.98	0.89
Reliability of contents	0.89	0.89	0.78	0.69	0.93
Completeness of resources	0.79	0.91	0.87	0.91	0.96
Data Security	0.76	0.67	0.88	0.95	0.89

TABLE II. PARTICIPANTS' EVALUATION ON LEARNING / CBT DESIGN

Features	Participants' Scores		
	σ	μ	+Di
Design Effectiveness	0.27	0.94	0.89
Availability of Video Conference	0.27	0.87	0.89
CBT Integration	0.23	0.82	0.73
Technical Support	0.33	0.90	0.95
Effectives Use by Teacher/Learner	0.28	0.81	0.78
Duration and Reciprocity	0.38	0.81	0.72
Reliability of contents	0.31	0.80	0.86
Completeness of resources	0.29	0.93	0.76
Data Security	0.13	0.85	0.72

Table II shows participant evaluation with mean, standard deviation and dyadic interaction between participants and the system. It yields a mean effectiveness in the ensemble use with over 90%, and a dyadic interaction that is above 70%. Also the relationship on the interaction between the participants and the e-learning platform cum system is sufficient.

B. Result Findings

The resulting resumption capability allows student's time to be recovered soon as fault is rectified. The system records the last activity during connectivity loss, and restores the time spent on it. If loss exceeds a threshold of 500 seconds, time is restored only if it persists. Ensemble resumes activities via automatic retrieval of remainder time, question-set and selected choices as in Fig. 2. The IP address and hostname of the client system, assigned to the user's registration number during the exam, shall also be used to restart the learner's session from the same device if and when the power is lost. This agrees with [75], [76].

Furthermore, the complexity with which examinations are scheduled is eliminated by using a series of simultaneous tests. The database is filled with a list of registered courses for each student. A student is entitled to write any scheduled course(s) registered, once the course schedule has been set. An algorithm for random selection of questions, is implemented

to avoid the same sequence of questions and answers being assigned to students. Eq. (2) is used to validate the algorithm so that the probability a student gets the same sequence of questions with the same sequence of choices is denoted by $P(R)$:

$$P(R) = \frac{Q \cdot S}{N \cdot Q! \cdot \pi_1^Q C_i!} \quad (2)$$

Where Q is the number of exam's question is Q , C_i is the number of choices for Q_i (where $i=1...Q$), the number of students is S , and the total number of questions stored in the database for this exam is N (where $Q \leq N$).

C. Discussion of Findings

We model the AdaCoBaTe ensemble to resolve the issues of fault-tolerance and resumption capability to reduce the effect of the occurrence of faults that may occur from power, network and physical infrastructure failure(s) – that may disrupt or terminate an examination. This can result in timing out that will eventually supposedly log out a user. To restore lost time, the system would monitor client computer connectivity to a server in order to solve this problem. System will restore the clock to when a student has last been in touch with this server if connectivity is lost for more than 500 seconds. In addition to restoring the rest of the period, if an exam is restarted, question set and selected choices are restored. This agrees with [77]–[79].

The algorithm for random selection of questions, its random distribution and random choice of responses – were developed and used. In general, it reduced examination malpractice as is corroborated by Fig. 2. A database is kept of the randomized question sets and selected answers, which are assigned to students when examinations resume. The random algorithm uses inbuilt-PHP shuffle function to ensure that sets of randomly selected questions are distributed normally. A comparison of the features of the proposed system to some popular online testing systems has been carried out. Automated resumption of tests will facilitate the adoption of Web based examinations by institutions that do not require uninterrupted power or Internet connection. This agrees with [80]–[82].

Further research can be poised toward the application of this study to online exams with descriptive questions, and not just multi-choice questions, and seek new means to track all assigned questions (alongside the selected choices as database size grows larger and for larger examinations). Also, to advance its security also, further studies can propose live image capture for students undertaking the exams as a means to detect impersonation when the exam is strictly online/web-enabled.

IV. CONCLUSION

To reposition education as a key integral facet of the society as recovery strategies against the impacts of covid-19 pandemic – requires strong policies, which will yield unexpected high-end results. And though, the impact in Nigeria (on a grander-scale) was not as projected – the shocks and disruptions as experienced with covid-19, still raises important educational concerns that only new reforms can

answer and help with national recovery. The questions raised as reflection of the local realities vis-à-vis implementation of platforms to exchange experiences will in turn, foster effective strategies cum policies to help repair the wreckage impacted on the society by covid-19 – and mitigate future pandemic. This CBT ensemble can be implemented on a Campus-Intranet design to help curb the issues of privileged control as well as security risks that comes with its access via the Internet. This intranet-based design mode will also curb the issues of interoperability. The personalized delivery outcome on the e-learning setting seeks to minimize the effects of anonymity and social presence gaps, advances an improved learner satisfaction with the tailored content delivery, improved cognition, and better engaged participation in class of learners.

REFERENCES

- [1] O. Eboka and A. A. Ojugo, "Mitigating technical challenges via redesigning campus network for greater efficiency, scalability and robustness: A logical view," *Int. J. Mod. Educ. Comput. Sci.*, vol. 12, no. 6, pp. 29–45, 2020, doi: 10.5815/ijmecs.2020.06.03.
- [2] G. Ocak and A. Yamaç, "Examination of the relationships between fifth graders' self-regulated learning strategies, motivational beliefs, attitudes, and achievement," *Educ. Sci. Theory Pract.*, vol. 13, no. 1, pp. 380–387, 2013.
- [3] J. R. Fraenkel, N. E. Wallen, and H. H. Hyun, *How to design and evaluate research in education*. New York, USA: McGraw-Hill, 2012.
- [4] F. Reichert, D. Lange, and L. Chow, "Educational beliefs matter for classroom instruction: A comparative analysis of teachers' beliefs about the aims of civic education," *Teach. Teach. Educ.*, vol. 98, pp. 1–13, 2020, doi: 10.1016/j.tate.2020.103248.
- [5] E. H. Mahvelati, "Learners' perceptions and performance under peer versus teacher corrective feedback conditions," *Stud. Educ. Eval.*, vol. 70, 2021, doi: 10.1016/j.stueduc.2021.100995.
- [6] S. R. Guntur, R. R. Gorrepati, and V. R. Dirisala, "Internet of Medical Things," in *Medical Big Data and Internet of Medical Things*, no. October 2018, Boca Raton: Taylor & Francis, [2019]: CRC Press, 2018, pp. 271–297. doi: 10.1201/9781351030380-11.
- [7] J. H. Joloudari, R. Alizadehsani, and I. Nodehi, "Resource allocation optimization using artificial intelligence methods in various computing paradigms: A Review," no. March, 2022, doi: 10.13140/RG.2.2.32857.39522.
- [8] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. C. Odiakaose, F. U. Emordi, and A. A. Ojugo, "DeLClustE: Protecting Users from Credit-Card Fraud Transaction via the Deep-Learning Cluster Ensemble," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 94–100, 2023, doi: 10.14569/IJACSA.2023.0140610.
- [9] A. M. Flanagan, D. C. Cormier, and O. Bulut, "Achievement may be rooted in teacher expectations: examining the differential influences of ethnicity, years of teaching, and classroom behaviour," *Soc. Psychol. Educ.*, vol. 23, pp. 1429–1448, 2020, doi: 10.1007/s11218-020-09590-y.
- [10] D. N. D. N. Roberson, S. Zach, N. Chores, and I. Rosenthal, "Self Directed Learning: A Longstanding Tool for Uncertain Times," *Creat. Educ.*, vol. 12, no. 05, pp. 1011–1026, 2021, doi: 10.4236/ce.2021.125074.
- [11] R. G. Brockett and R. Hiemstra, *Self-Direction in Adult Learning*. London and New York: Routledge, 2018. doi: 10.4324/9780429457319.
- [12] W. H. Goodridge, O. Lawanto, and H. B. Santoso, "A Learning Style Comparison between Synchronous Online and Face-to-Face Engineering Graphics Instruction," *Int. J. Mod. Educ. Comput. Sci.*, vol. 15, no. 2, pp. 1–14, 2017, doi: 10.5539/ies.v10n2p1.
- [13] M. I. Akazue, A. A. Ojugo, R. E. Yoro, B. O. Malasowe, and O. Nwankwo, "Empirical evidence of phishing menace among undergraduate smartphone users in selected universities in Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 3, pp. 1756–1765, Dec. 2022, doi: 10.11591/ijeecs.v28.i3.pp1756-1765.
- [14] M. I. Akazue, R. E. Yoro, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo, "Improved services traceability and management of a food value chain using block-chain network: a case of Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 3, pp. 1623–1633, 2023, doi: 10.11591/ijeecs.v29.i3.pp1623-1633.
- [15] A. A. Ojugo and R. E. Yoro, "Extending the three-tier constructivist learning model for alternative delivery: ahead the COVID-19 pandemic in Nigeria," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 3, p. 1673, Mar. 2021, doi: 10.11591/ijeecs.v21.i3.pp1673-1682.
- [16] O. S. Adewale, E. O. Ibam, and B. K. Alese, "A web-based virtual classroom system model," *Turkish Online J. Distance Educ.*, vol. 13, no. 1, pp. 211–223, 2012.
- [17] R. Amelia, G. Kadarisma, N. Fitriani, and Y. Ahmadi, "The effect of online mathematics learning on junior high school mathematic resilience during covid-19 pandemic," *J. Phys. Conf. Ser.*, vol. 1657, no. 1, p. 012011, Oct. 2020, doi: 10.1088/1742-6596/1657/1/012011.
- [18] M. Belot and D. Webbink, "Do Teacher Strikes Harm Educational Attainment of Students?," *LABOUR*, vol. 24, no. 4, pp. 391–406, Dec. 2010, doi: 10.1111/j.1467-9914.2010.00494.x.
- [19] R. E. Yoro, F. O. Aghware, B. O. Malasowe, O. Nwankwo, and A. A. Ojugo, "Assessing contributor features to phishing susceptibility amongst students of petroleum resources varsity in Nigeria," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1922, Apr. 2023, doi: 10.11591/ijeecs.v13i2.pp1922-1931.
- [20] F. Borgonovi and A. Ferrara, "A Longitudinal Perspective on the Effects of COVID-19 on Students' Resilience. The Effect of the Pandemic on the Reading and Mathematics Achievement of 8th and 5th Graders in Italy," *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4025865.
- [21] F. Agostinelli, M. Doepke, G. Sorrenti, and F. Zilibotti, "When the great equalizer shuts down: Schools, peers, and parents in pandemic times," *J. Public Econ.*, vol. 206, p. 104574, Feb. 2022, doi: 10.1016/j.jpubeco.2021.104574.
- [22] A. A. Ojugo and O. Nwankwo, "Spectral-Cluster Solution For Credit-Card Fraud Detection Using A Genetic Algorithm Trained Modular Deep Learning Neural Network," *JINAV J. Inf. Vis.*, vol. 2, no. 1, pp. 15–24, Jan. 2021, doi: 10.35877/454RI.jinav274.
- [23] H. Patrinos, E. Vegas, and R. Carter-Rau, "An Analysis of COVID-19 Student Learning Loss," *Educ. Glob. Pract. Policy Res. Work. Pap.* 10033, vol. 10033, no. May, pp. 1–31, 2022, doi: 10.1596/1813-9450-10033.
- [24] M. Brindlmayer, R. Khadduri, A. Osborne, A. Briansó, and E. Cupito, "Prioritizing learning during covid-19: The Most Effective Ways to Keep Children Learning During and Post-Pandemic," *Glob. Educ. Evid. Advis. Panel*, no. January, pp. 1–21, 2022.
- [25] R. E. Yoro, F. O. Aghware, M. I. Akazue, A. E. Ibor, and A. A. Ojugo, "Evidence of personality traits on phishing attack menace among selected university undergraduates in Nigerian," *Int. J. Electr. Comput. Eng.*, vol. 13, no. 2, p. 1943, Apr. 2023, doi: 10.11591/ijeecs.v13i2.pp1943-1953.
- [26] M. Dominic and S. Francis, "An Adaptable E-Learning Architecture Based on Learners' Profiling," *Int. J. Mod. Educ. Comput. Sci.*, vol. 7, no. 3, pp. 26–31, 2015, doi: 10.5815/ijmecs.2015.03.04.
- [27] E. D. Ananga, "Gender Responsive Pedagogy for Teaching and Learning: The Practice in Ghana's Initial Teacher Education Programme," *Creat. Educ.*, vol. 12, no. 04, pp. 848–864, 2021, doi: 10.4236/ce.2021.124061.
- [28] A. A. Ojugo and A. O. Eboka, "Assessing Users Satisfaction and Experience on Academic Websites: A Case of Selected Nigerian Universities Websites," *Int. J. Inf. Technol. Comput. Sci.*, vol. 10, no. 10, pp. 53–61, 2018, doi: 10.5815/ijitcs.2018.10.07.
- [29] A. A. Ojugo and D. O. Otakore, "Redesigning Academic Website for Better Visibility and Footprint: A Case of the Federal University of Petroleum Resources Effurun Website," *Netw. Commun. Technol.*, vol. 3, no. 1, p. 33, Jul. 2018, doi: 10.5539/nct.v3n1p33.
- [30] F. S. Hilyana, "Implementation of Schoology-based E-Learning to Improve the ANEKA-based Character," 2018. doi: 10.4108/eai.24-10-2018.2280558.
- [31] J. M. Jiménez-Olmedo, A. Penichet-Tomás, B. Pucio, and S. Sebastián-Amat, "Comparative Analysis of Content Learning Through Schoology

- and Micro-Teaching in Higher Education,” *EDULEARN18 Proc.*, vol. 1, no. July, pp. 6348–6352, 2018, doi: 10.21125/edulearn.2018.1507.
- [32] C. E. Eko, I. E. Eteng, and E. E. Essien, “Design and Implementation of a Fault Tolerant Web-Based Examination System for Developing Countries,” *Eastern-European J. Enterp. Technol.*, vol. 1, no. 2–115, pp. 58–67, 2022, doi: 10.15587/1729-4061.2022.253146.
- [33] A. A. Ojugo, E. Ugboh, C. C. Onochie, A. O. Eboka, M. O. Yerokun, and I. J. B. Iyawa, “Effects of Formative Test and Attitudinal Types on Students’ Achievement in Mathematics in Nigeria,” *African Educ. Res. J.*, vol. 1, no. 2, pp. 113–117, 2013, [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1216962&site=ehost-live>
- [34] S. D. Durojaye, E. O. Okon, and D. D. Samson, “Software Quality and Usability for Computer-Based Test in Tertiary Institution in Nigeria: A Case Study of Kogi State University,” *Am. J. Educ. Res.*, vol. 3, no. 10, pp. 1224–1229, 2015, doi: 10.12691/education-3-10-3.
- [35] T. M. Fagbola, A. A. Adigun, and A. O. Oke, “Computer-Based Test (Cbt) System For University Academic Enterprise Examination,” *Int. J. Sci. Technol. Res.*, vol. 2, no. 8, pp. 336–342, 2013, [Online]. Available: <http://www.ijstr.org/final-print/aug2013/Computer-based-Test-Cbt-System-For-University-Academic-Enterprise-Examination.pdf>
- [36] G. F. Fragulis, M. Papatsimouli, L. Lazaridis, and I. A. Skordas, “An Online Dynamic Examination System (ODES) based on open source software tools,” *Softw. Impacts*, vol. 7, p. 100046, 2021, doi: 10.1016/j.simpa.2020.100046.
- [37] G. F. Fragulis, L. Lazaridis, M. Papatsimouli, and I. A. Skordas, “O.D.E.S.: An Online Dynamic Examination System based on a CMS Wordpress plugin,” *South-East Eur. Des. Autom. Comput. Eng. Comput. Networks Soc. Media Conf. SEEDA_CECNSM 2018*, no. May, 2018, doi: 10.23919/SEEDA-CECNSM.2018.8544928.
- [38] A. A. Ojugo and R. E. Yoro, “Forging a deep learning neural network intrusion detection framework to curb the distributed denial of service attack,” *Int. J. Electr. Comput. Eng.*, vol. 11, no. 2, pp. 1498–1509, 2021, doi: 10.11591/ijece.v11i2.pp1498-1509.
- [39] O. Adebayo and S. M. Abdulhamid, “E- Exams System for Nigerian Universities with Emphasis on Security and Result Integrity,” *Int. J. Comput. Internet Manag.*, vol. 18, no. 2, pp. 47.1–47.13, 2019, [Online]. Available: <http://arxiv.org/abs/1402.0921>
- [40] M. Ajinaja, “The Design and Implementation of a Computer Based Testing System Using Component-Based Software Engineering,” *Int. J. Comput. Sci. Technol.*, vol. 8, no. 1, pp. 58–65, 2017.
- [41] L. O. Akazua, K. S. Nwizege, F. O. Philip-Kpae, J. Danamina, B. G. Akoba, and P. G. Irimiagha, “Positive Impacts of Online Examination with Answer-Correction Feedback in Nigeria,” *Int. J. Emerg. Sci. Eng.*, vol. 4, no. 7, pp. 1–12, 2016.
- [42] H. Albazar, “A New Automated Forms Generation Algorithm for Online Assessment,” *J. Inf. Knowl. Manag.*, vol. 19, no. 01, p. 2040008, Mar. 2020, doi: 10.1142/S0219649220400080.
- [43] L. B. Barik, B. Patel, and A. Barik, “You are watching system (YAW System): An Agent based secure real-time monitoring, capturing and analysis system for client activities on the Expert Agent’s Screen,” *ACM Int. Conf. Proceeding Ser.*, no. January, pp. 507–510, 2011, doi: 10.1145/1947940.1948045.
- [44] A. A. Ojugo, A. O. Eboka, R. E. Yoro, M. O. Yerokun, and F. N. Efozia, “Hybrid model for early diabetes diagnosis,” in *2015 Second International Conference on Mathematics and Computers in Sciences and in Industry (MCSI)*, 2015, pp. 55–65. doi: 10.1109/MCSI.2015.35.
- [45] [45] R. O. Bello, M. Olugbebi, A. O. Babatunde, B. O. Bello, and S. I. Bello, “A University Examination Web Application Based on Linear-Sequential Life Cycle Model,” *Daffodil Int. Univ. J. Sci. Technol.*, vol. 12, no. 1, p. 25, 2017.
- [46] O. B. Chibuzo and D. O. Isiaka, “Design and Implementation of Secure Browser for Computer-Based Tests,” *Int. J. Innov. Sci. Res. Technol.*, vol. 5, no. 8, pp. 1347–1356, 2020, doi: 10.38124/ijisrt20aug526.
- [47] R. Bello, M. Olugbebi, A. Babatunde, B. Bello, and S. Bello, “Design and Implementation of Web- based Examination system for the University,” *J. Comput. Sci. Control Sci.*, vol. 9, no. 2, pp. 5–9, 2016.
- [48] A. A. Ojugo, P. O. Ejeh, C. C. Odiakaose, A. O. Eboka, and F. U. Emordi, “Improved distribution and food safety for beef processing and management using a blockchain-tracer support framework,” *Int. J. Informatics Commun. Technol.*, vol. 12, no. 3, p. 205, Dec. 2023, doi: 10.11591/ijict.v12i3.pp205-213.
- [49] H. Danladi and A. K. Dodo, “A comparative Analysis of Joint Admissions and Matriculation Board’s (JAMB) Performance, Pre and Post Electronic migration,” *Int. J. Humanit. Soc. Sci.*, vol. 6, no. 5, pp. 80–85, 2019, doi: 10.14445/23942703/ijhss-v6i5p111.
- [50] F. O. Aghware, R. E. Yoro, P. O. Ejeh, C. Odiakaose, F. U. Emordi, and A. A. Ojugo, “Sentiment Analysis in Detecting Sophistication and Degradation Cues in Malicious Web Contents,” *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 01, pp. 653–665, 2023.
- [51] A. A. Ojugo, M. I. Akazue, P. O. Ejeh, C. Odiakaose, and F. U. Emordi, “DeGATraMoNN : Deep Learning Memetic Ensemble to Detect Spam Threats via a Content-Based Processing,” *Kongzhi yu Juece/Control Decis.*, vol. 38, no. 01, pp. 667–678, 2023.
- [52] S. Iskandarov, “Develop a centralized and secure online testing system for a large number of users,” *J. Inf. Knowl. Manag.*, vol. 23, no. October, pp. 1–6, 2020.
- [53] D. A. Oyemade, R. J. Ureigho, F. . Imoukhome, E. U. Omoregbee, J. Akpojaro, and A. A. Ojugo, “A Three Tier Learning Model for Universities in Nigeria,” *J. Technol. Soc.*, vol. 12, no. 2, pp. 9–20, 2016, doi: 10.18848/2381-9251/CGP/v12i02/9-20.
- [54] M. I. Younis, M. S. Hussein, and M. I. Younis, “Construction of an Online Examination System with Resumption and Randomization Capabilities,” *Int. J. Comput. Acad. Res.*, vol. 4, no. 2, pp. 62–82, 2015, [Online]. Available: <http://www.meacse.org/ijcar>
- [55] D. L. Chen, S. Ertac, T. Evgeniou, X. Miao, A. Nadaf, and E. Yilmaz, “Grit and Academic Resilience During the Covid-19 Pandemic,” *SSRN Electron. J.*, 2022, doi: 10.2139/ssrn.4001431.
- [56] J. Crawford, K. Butler-Henderson, and J. Rudolph, “COVID-19: 20 countries’ higher education intra-period digital pedagogy responses,” *J. Appl. Learn. Teach.*, vol. 3, no. 1, Apr. 2020, doi: 10.37074/jalt.2020.3.1.7.
- [57] E. Haiping, N. Kadhila, and L. M. Josua, “Using Digital Technology in Transforming Assessment in Higher Education Institutions beyond COVID-19,” *Creat. Educ.*, vol. 13, no. 07, pp. 2157–2167, 2022, doi: 10.4236/ce.2022.137136.
- [58] A. A. Ojugo and A. O. Eboka, “Empirical Bayesian network to improve service delivery and performance dependability on a campus network,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 3, p. 623, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp623-635.
- [59] A. A. Ojugo and O. D. Otakore, “Intelligent cluster connectionist recommender system using implicit graph friendship algorithm for social networks,” *IAES Int. J. Artif. Intell.*, vol. 9, no. 3, p. 497–506, 2020, doi: 10.11591/ijai.v9.i3.pp497-506.
- [60] S. Fabriz, J. Mendzheritskaya, and S. Stehle, “Impact of Synchronous and Asynchronous Settings of Online Teaching and Learning in Higher Education on Students’ Learning Experience During COVID-19,” *Front. Psychol.*, vol. 12, no. October, pp. 1–16, Oct. 2021, doi: 10.3389/fpsyg.2021.733554.
- [61] N. F. Duarte Filho and E. F. Barbosa, “A contribution to the quality evaluation of mobile learning environments,” in *2013 IEEE Frontiers in Education Conference (FIE)*, Oct. 2013, pp. 379–382. doi: 10.1109/FIE.2013.6684851.
- [62] M. L. Fioravanti and E. F. Barbosa, “MLearning-PL: a pedagogical pattern language for mobile learning applications,” in *HILLSIDE Proc. of Conf. on Pattern Lang. of Prog.*, Oct. 2017, pp. 1–29. doi: 10.5753/cbie.wcbie.2018.64.
- [63] B. F. Komolafe, O. T. Fakayode, A. Osidipe, F. Zhang, and X. Qian, “Evaluation of Online Pedagogy among Higher Education International Students in China during the COVID-19 Outbreak,” *Creat. Educ.*, vol. 11, no. 11, pp. 2262–2279, 2020, doi: 10.4236/ce.2020.1111166.
- [64] A. A. Ojugo and D. A. Oyemade, “Boyer moore string-match framework for a hybrid short message service spam filtering technique,” *IAES Int. J. Artif. Intell.*, vol. 10, no. 3, pp. 519–527, 2021, doi: 10.11591/ijai.v10.i3.pp519-527.
- [65] A. Hidayat and V. G. Utomo, “Adaptive Online Module Prototype for Learning Unified Modelling Language (UML),” *Int. J. Electr. Comput.*

- Eng., vol. 6, no. 6, p. 2931, Dec. 2016, doi: 10.11591/ijece.v6i6.pp2931-2938.
- [66] V.-D. Nguyen, D.-N. Tran, H.-H. Tran, T.-N. Phan, T. Danh, and H.-N. Tran, "Blended Learning Model-Based Local Education for Vietnamese Primary School Students," *Rev. Int. Geogr. Educ.*, vol. 11, no. 8, pp. 1684–1694, 2022, doi: 10.48047/rigeo.11.08.145.
- [67] A. E. Ibor, E. B. Edim, and A. A. Ojugo, "Secure Health Information System with Blockchain Technology," *J. Niger. Soc. Phys. Sci.*, vol. 5, no. 992, pp. 1–8, 2023, doi: 10.46481/jnsps.2022.992.
- [68] G. Nguyen et al., "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.
- [69] A. M. Muxtorjonovna, "Significance Of Blended Learning In Education System," *Am. J. Soc. Sci. Educ. Innov.*, vol. 02, no. 08, pp. 507–511, 2020, doi: 10.37547/tajssei/volume02issue08-82.
- [70] D. G. H. Divayana, "Aneka-based asynchronous and synchronous learning design and its evaluation as efforts for improving cognitive ability and positive character of students," *Int. J. Mod. Educ. Comput. Sci.*, vol. 13, no. 5, pp. 14–22, 2021, doi: 10.5815/ijmecs.2021.05.02.
- [71] A. A. Ojugo, C. O. Obruché, and A. O. Eboka, "Empirical Evaluation for Intelligent Predictive Models in Prediction of Potential Cancer Problematic Cases In Nigeria," *ARRUS J. Math. Appl. Sci.*, vol. 1, no. 2, pp. 110–120, Nov. 2021, doi: 10.35877/mathscience614.
- [72] A. A. Ojugo, C. O. Obruché, and A. O. Eboka, "Quest For Convergence Solution Using Hybrid Genetic Algorithm Trained Neural Network Model For Metamorphic Malware Detection," *ARRUS J. Eng. Technol.*, vol. 2, no. 1, pp. 12–23, Nov. 2021, doi: 10.35877/jetech613.
- [73] A. A. Ojugo, A. O. Eboka, E. O. Okonta, R. E. Yoro, and F. O. Aghware, "Predicting Behavioural Evolution on a Graph-Based Model," *Adv. Networks*, vol. 3, no. 2, p. 8, 2015, doi: 10.11648/j.net.20150302.11.
- [74] A. R. Sunday, S. O. Etuk, and I. M. Kolo, "Biometry, Encryption and Spyware: A Multi-factor Security and Authentication Mechanism for JAMB E- Examination," *Int. J. Appl. Inf. Syst.*, vol. 12, no. 32, pp. 17–26, 2020.
- [75] N. B. Udoekanem, "Human Capacity Training for Security of Life , Property and Investment: A Challenge for Estate Management Education," *J. Sci. Technol. Math. Educ.*, vol. 7, no. 3, pp. 2–293, 2018.
- [76] A. Suleiman and N. Nachandiya, "Computer Based Testing (CBT) System for GST Exams in Adamawa State University, Mubi," *Asian J. Res. Comput. Sci.*, no. November, pp. 1–11, 2018, doi: 10.9734/ajrcos/2018/v2i1124776.
- [77] J. A. Abah, O. HONMANE, T. J. Age, and S. O. OGBULE, "Design of Single-User-Mode Computer-Based Examination System for Senior Secondary Schools in Onitsha North Local Government Area of Anambra State, Nigeria," *SSRN Electron. J.*, vol. 6, no. January, pp. 12–21, 2022, doi: 10.2139/ssrn.4061818.
- [78] S. Ejim, "Computer based examination system with multi-factor authentication and message notification features," Abubakar Tafawa Balewa University, Bauchi, 2017. doi: 10.13140/RG.2.2.14713.88167.
- [79] F. F. Haryani and D. Maryono, "Online learning in Indonesian higher education : New indicators during the COVID-19 pandemic," vol. 12, no. 3, pp. 1262–1270, 2023, doi: 10.11591/ijere.v12i3.24086.
- [80] I. O. Izu-Okpara, O. C. Nwokonkwo, and A. M. John-Otumu, "An Implementation of K-NN Classification Algorithm for Detecting Impersonators in Online Examination Environment," *J. Adv. Comput. Commun. Inf. Technol.*, vol. 1, no. April, pp. 8–17, 2021, doi: 10.37121/jaccit.v1.153.
- [81] O. Kainz, D. Cymbalák, and F. Jakab, "Adaptive Web-Based System for Examination with Cheating Prevention Mechanism," *Lect. Notes Softw. Eng.*, vol. 3, no. 2, pp. 90–94, 2015, doi: 10.7763/Inse.2015.v3.172.
- [82] L. Lazaridis, M. Papatsimouli, and G. F. Fragulis, "S.A.T.E.P.," in *Proceedings of the SouthEast European Design Automation, Computer Engineering, Computer Networks and Social Media Conference*, Sep. 2016, pp. 92–97. doi: 10.1145/2984393.2984395.

A Yolo-based Violence Detection Method in IoT Surveillance Systems

Hui Gao

College of Computer and Information Engineering
Xinxiang University
Xinxiang 453000, Henan, China

Abstract—Violence detection in Internet of Things (IoT)-based surveillance systems has become a critical research area due to their potential to provide early warnings and enhance public safety. There have been many types of research on vision-based systems for violence detection, including traditional and deep learning-based methods. Deep learning-based methods have shown great promise in ameliorating the efficiency and accuracy of violence detection. Despite the recent advances in violence detection using deep learning-based methods, significant limitations and research challenges still need to be addressed, including the development of standardized datasets and real-time processing. This study presents a deep learning method based on You Only Look Once (YOLO) algorithm for the violence detection task to overcome these issues. We generate a model for violence detection using violence and non-violence images in a prepared dataset divided into testing, validation, and training sets. Based on accepted performance indicators, the produced model is assessed. The experimental results and performance evaluation show that the method accurately identifies violence and non-violence classes in real-time.

Keywords—Violence detection; IoT; surveillance systems; Yolo; deep learning

I. INTRODUCTION

The use of Internet of Things (IoT) based surveillance systems has elevated significantly in the past few years, particularly for detecting and preventing violent incidents in public spaces[1–3]. Violence detection in IoT-based surveillance systems has become a critical research area due to its potential to provide early warnings and enhance public safety[4,5]. These systems can process and analyze real-time data from sensors and cameras, enabling quick and efficient identification of violent incidents [6].

There have been significant advances in violence detection technologies in IoT-based surveillance systems in recent years [7,8]. There has been significant research on vision-based systems for violence detection, including traditional and deep learning-based methods [9,10]. Traditional methods, such as motion detection, background subtraction, and object tracking, have been widely used for violence detection in surveillance systems [6,11,12]. However, these methods have limitations in terms of accuracy and robustness, particularly in complex and cluttered environments.

Recent studies have shown that deep learning-based model, like recurrent neural networks (RNNs), convolutional neural networks (CNNs), and YOLO, is able to significantly improve

the accuracy and efficiency of violence detection in vision-based systems [13–15]. These models can process and analyze image and video data, extract complex features, and identify violent events in real time.

Despite the recent advances in violence detection, using deep learning-based methods has shown great promise in ameliorating the efficiency and accuracy of violence detection [13,16]. Nevertheless, significant limitations and research challenges still need to be addressed, including the development of standardized datasets and real-time processing algorithms. According to these challenges, the lack of standardized datasets leads to generating inaccurate model for violence detection. Moreover, it is required to address an efficient model to perform in real-time requirement. This makes comparing the different models' performances challenging and limits their generalizability. Addressing this challenge is essential to advance the field and ensure the accurate and efficient detection of violent events in surveillance systems.

To deal with the research challenge in this work, the YOLO algorithm is utilized for the violence detection task in order to overcome these issues. The most recent object identification technique, YOLO, is highly accurate and quick in detecting several items in a picture. We generate a model for violence detection using violence and non-violence images in a dataset that has been divided into testing, validation, and training sets. Based on accepted performance indicators, the produced model is assessed. The system can be taught to recognize violence patterns and accurately identify violence and non-violence classes in real time.

The rest of this paper is structured as follows; Section II presents literature review. Section III discuss about the methodology. Experimental results and performance evaluation presents in Section IV Finally, the paper concludes in Section V.

II. LITERATURE REVIEW

This section presents the literature review and related works on the violence detection research domain. Ullah et al. [3], in IoT-based industrial surveillance networks, this research presented an edge vision technique with AI assistance for violence detection. The technique uses cloud computing, edge devices, and deep learning-based algorithms to analyse video data and identify possible real-time risks. Some important features are custom datasets for training, cloud computing and

edge device integration, and real-time notifications for possible risks. The method successfully detects violent occurrences with low false-positive rates and high accuracy. The method's drawbacks include the sizeable computational resources needed for in-the-moment data processing and analysis, as well as the hefty infrastructure expenditures.

AIDahoul et al. [17] rendered a method for violence detection utilizing a Convolutional Neural Network-Long Short Term Memory (CNN-LSTM) based IoT node. The suggested method utilizes a custom dataset for training the CNN-LSTM model, which analyzes the video data captured by the IoT node to detect violent events. The system can process and analyze data in real-time, quickly detecting potential threats. The key features of the proposed method include the use of a CNN-LSTM model for violence detection, the integration of IoT devices for data capture, and the ability to perform real-time analysis of data. The study found that the proposed approach achieved high levels of accuracy in detecting violent events with low false-positive rates. One limitation of the proposed approach is the potential for high power consumption by the IoT node when processing and analyzing data. The authors also note that the performance of the system may vary based on environmental factors and the specific application scenario.

In [18], the research presented a weakly supervised method for detecting violence in surveillance footage. In order to identify probable violent occurrences without needing manual annotation of the training data, the technique employs a Convolutional Neural Network (CNN) model to categorise video frames as violent or non-violent. The approach's main characteristics are using CNN models for classification and weakly supervised learning, eliminating the requirement for annotated training data. The study discovered that the proposed strategy outperformed earlier state-of-the-art approaches in achieving high accuracy in identifying violent incidents. The approach may still need some manual annotation, the authors point out, in order to operate at its best.

Abdali et al. [19] developed a CNN and Long Short-Term Memory (LSTM) model-based real-time violence detection method. To analyse video frames in real time and pinpoint violent situations, the proposed technique combines the advantages of CNN and LSTM. The approach's primary characteristics include real-time video processing, a bespoke training dataset, and highly accurate violent event detection with low false-positive rates. According to the study, the suggested solution beats current approaches in terms of processing speed and accuracy, making it appropriate for use in practical applications. The approach could involve a lot of computational power, and further study is needed to improve it for various settings and environments.

III. METHODOLOGY

This section discusses the details of the procedures in our methodology. This method consists of dataset description, dataset set preparation, Yolo algorithm, and training of the Yolo model. The corresponding details explain in the following sections.

A. Description of the Dataset

The dataset includes 3333 images of resolution 416 x 416 pixels with the annotated objects. The annotations include bounding boxes around people and objects of interest and labels indicating whether the object is associated with violent behavior. The dataset includes examples of different types of violence, including fights, weapons, and attacks. The dataset is intended for use with the YOLO (You Only Look Once) algorithm, a popular object detection algorithm known for its speed and accuracy.

B. Dataset Preparation

This study's provided dataset for violence detection has undergone several pre-processing and augmentation procedures. Pre-processing refers to the process of preparing the data for machine learning tasks. In this dataset, pre-processing included resizing all images to a resolution of 416 x 416 pixels, the input size required by the YOLO algorithm. Additionally, the dataset was converted to the YOLO format, which involves creating text files that contain the bounding box annotations and labels for each image.

Augmentation procedures are utilized to enhance the dataset's diversity and size artificially, improving the model's performance by making it more robust to variations in the input data. The dataset was augmented using various techniques, such as random scaling, random horizontal flipping, random rotation, and random translation. These techniques were applied to each image and corresponding annotations to create new training samples with slightly different characteristics.

Random horizontal flipping involves randomly flipping each image horizontally, which increases the diversity of the dataset and helps prevent overfitting. Random scaling involves randomly scaling each image by a factor of 0.25 to 2.0, which helps the model learn to recognize objects at different scales. Random translation involved randomly shifting each image horizontally and vertically by up to 20% of its width and height, respectively, which helps the model learn to recognize objects in different positions. Random rotation involves randomly rotating each image by up to 10 degrees, which helps the model learn to recognize objects from different angles.

All of these pre-processing and augmentation procedures were performed to enhance the diversity and quality of the dataset, which is able to lead to better performance and generalization of the violence detection model.

C. YOLO Algorithm

YOLO (You Only Look Once) is an object detection algorithm that simultaneously forecasts class probabilities and bounding boxes for objects in an input image. It is a popular algorithm due to its real-time detection capabilities and high accuracy. The YOLO algorithm consists of two main components: a post-processing algorithm and a convolutional neural network (CNN). The CNN takes an input image and outputs a set of bounding boxes along with their class probabilities. The post-processing algorithm selects the most probable bounding boxes and discards the others.

YOLO has several versions, with YOLOv5 being the latest and most advanced version. YOLOv5 is faster and more

accurate than previous versions, thanks to several improvements, including using a backbone network architecture, improved feature extraction, and a more efficient post-processing algorithm. Fig. 1 demonstrates the architecture of YOLOv5 network. The backbone of YOLOv5 is called CSPDarknet, which stands for Cross Stage Partial Network. It is a modified version of the Darknet architecture used in previous versions of YOLO. CSPDarknet is composed of a series of convolutional layers that extract features from the input image. It is designed to be computationally efficient while still producing high-quality feature maps.

The neck of YOLOv5 is called PANet, which stands for Path Aggregation Network. It is a feature fusion module that combines features from different scales and resolutions. The PANet module uses a top-down pathway to aggregate features from high-resolution feature maps and a bottom-up pathway to aggregate features from low-resolution feature maps. The resulting feature maps are then fused to form a single feature map with rich information from multiple scales.

The head of YOLOv5 is called YOLOLayer. It is responsible for predicting class probabilities and bounding boxes for objects in the input image. YOLOLayer uses anchor boxes to predict the location and size of objects in the image. It also uses a softmax function to forecast the probability of each object belonging to a particular class. The final output of the YOLOLayer is a bounding box set with associated class probabilities.

The procedure in the YOLOv5 algorithm is based on the following steps:

1) *Input image*: YOLO takes an input image of size (width, height, channels) and resizes it to a fixed size (416x416x3) before feeding it to the network.

2) *CNN architecture*: The CNN architecture of YOLOv5 consists of a backbone network and several detection heads. The backbone network is based on CSPDarknet53, a variant of Darknet53. The detection heads are responsible for predicting bounding boxes and class probabilities.

3) *Feature extraction*: The feature extraction process is carried out by the backbone network, which generates feature maps of various resolutions. The feature maps are then fed to the detection heads.

4) *Bounding box prediction*: The detection heads predict a bounding box set for each feature map. The bounding boxes are represented as (x, y, w, h), where (x, y) is the center of the box, w is the width, and h is the height.

5) *Class probability prediction*: The detection heads also predict class probabilities for each bounding box. The class probabilities represent the probability that the object inside the bounding box belongs to a particular class.

6) *Non-maximum suppression*: After predicting bounding boxes and class probabilities, YOLOv5 applies non-maximum suppression to eliminate duplicate detections. Non-maximum suppression selects each object's most probable bounding box and discards the others.

In summary, YOLOv5 is a state-of-the-art object detection algorithm that utilizes a CNN to forecast class probabilities and bounding boxes for objects in an image. It has several improvements over previous versions, making it faster and more accurate. The algorithm consists of a backbone network, detection heads, and a post-processing algorithm that applies non-maximum suppression.

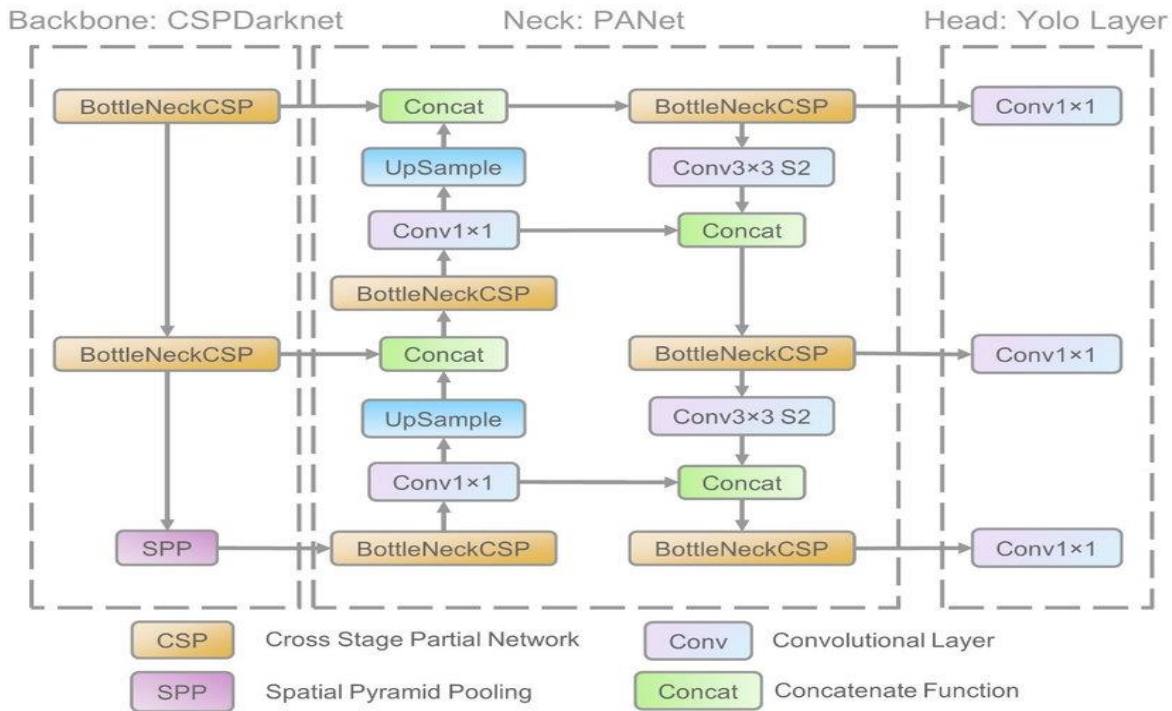


Fig. 1. The architecture of Yolov5 network [20].

D. Training the YOLOv5 Model

Training a YOLO (You Only Look Once) model for violence detection involves preparing the dataset, selecting a pre-trained model, fine-tuning it, and evaluating its performance. The first step discussed in section 3.2 is to prepare the dataset by resizing the images, annotating them with bounding boxes around the violent objects, and saving the annotations in a format the YOLO model can understand. Next, a suitable pre-trained YOLO model for violence detection must be selected for YOLOv5.

In the fine-tuning step, the pre-trained YOLO model is trained on the dataset, and its performance is assessed on the validation set. Hyperparameter tuning is able to be carried out to optimize the performance of the model. The batch size, learning rate, and number of epochs are some of the hyperparameters that can be tuned. Once the model is fine-tuned, its performance is assessed on the test set utilizing metrics such as F1 score, recall, and precision. If the model's performance is unsatisfactory, additional violent images can be added to the dataset, or the hyperparameters can be further tuned.

For training purpose, the dataset consisting of 2300 training samples, 662 testing samples, and 334 validation samples serves as the foundation for training the model YOLO model for violence detection using the given dataset of 2300 training samples, 662 testing samples, and 334 validation samples. Table I shows the proportion of each training, validation and testing sets. In the initial stage, the dataset is preprocessed to ensure consistency and compatibility with the YOLOv5 architecture. This involves resizing all images to a uniform input size, often in the form of squares, to facilitate streamlined processing. Annotations for each image are also processed to provide accurate bounding box coordinates and class labels corresponding to violent actions. These annotations are crucial for training the model to recognize and classify violence instances. The YOLOv5 model is then initialized with pre-trained weights, typically on a large-scale dataset, leveraging knowledge learned from a broad range of objects and features. Fine-tuning is performed on the violence detection dataset, allowing the model to adapt its features and parameters specifically for identifying violent actions. During training, the model iteratively adjusts its parameters by comparing predicted bounding boxes and class probabilities to the ground-truth annotations. This optimization process, often implemented using techniques like stochastic gradient descent, seeks to minimize the disparity between predictions and actual annotations.

To ensure the model generalizes well to new, unseen data, the training process employs techniques such as data augmentation. This involves applying transformations to the images, such as rotations, flips, and color variations, to expose the model to diverse scenarios it may encounter in real-world surveillance situations. Additionally, the training dataset is shuffled to prevent the model from memorizing the order of

samples. Throughout training, the model's performance is regularly evaluated using the validation dataset. Metrics like mean average precision (mAP) are calculated to assess the model's ability to precisely localize and classify violent actions. Training continues until the model's performance plateaus or shows satisfactory convergence.

TABLE I. POTATION TESTING, VALIDATION, AND TRAINING SETS

Set name	No. of samples	Set proportion (%)
<i>Training</i>	2300	70%
<i>Validation</i>	662	20%
<i>Testing</i>	334	10%

IV. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

The experimental findings and performance assessment of the suggested Yolov5 for violence detection on customized datasets are presented in this part: one research used to fall, no-fall, and half-classless films in a bespoke dataset. Utilizing various input image sizes, training datasets, and object detection thresholds, the study assessed Yolov5's performance. Fig. 2 demonstrates some examples of experimental results.

The experimental results and model performance evaluation for the YOLOv5 model for violence detection can be measured by various metrics such as Mean Average Precision (mAP), recall, and precision. Precision is the ratio of true positive detections to all of the model's positive detections. The formula for precision is:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

The ratio of precise positive detections to all positive cases in the dataset determines recall. The recall is the proportion of real positives to all real positives in the dataset. It displays the capacity of model to detect positive samples reliably. The formula for the recall is,

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

In precision and recall, where TP is the number of true positives (correctly detected violence), FP is the number of false positives (non-violences detected as anomalies), and FN is the number of false negatives (violence not detected by the algorithm). Based on obtained results, Table 2 presents performance measurements for the average precision rate for each class.

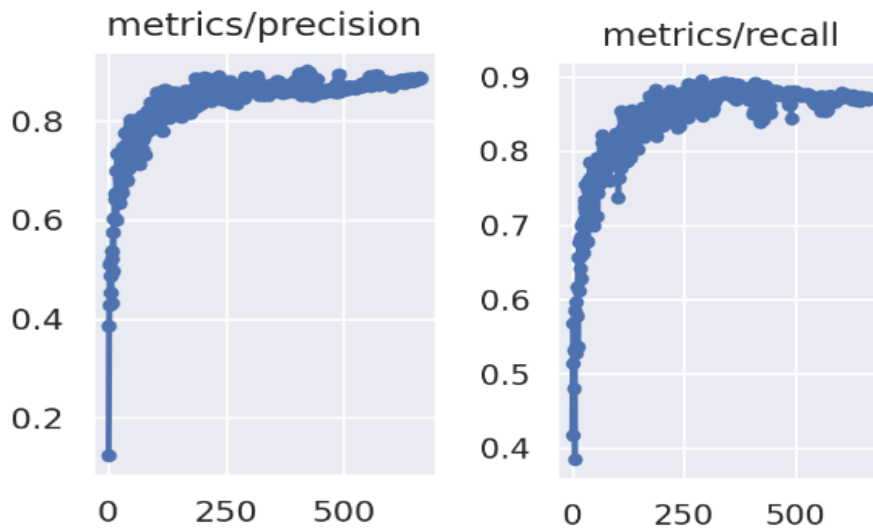
A popular performance metric, the mAP, accounts for memory and accuracy at different confidence levels. It returns a single scalar number summarising the model's overall detection performance as the average accuracy over a range of recall values. Fig. 3 illustrates performance metrics for generated Yolov5-based violence detection model.



Fig. 2. Experimental results.

TABLE II. AVERAGE PRECISION RATES BY CLASSES

Set name	Validation set	Testing set
<i>Violence</i>	93%	91%
<i>Non-Violence</i>	89%	90%
<i>All</i>	91%	91%



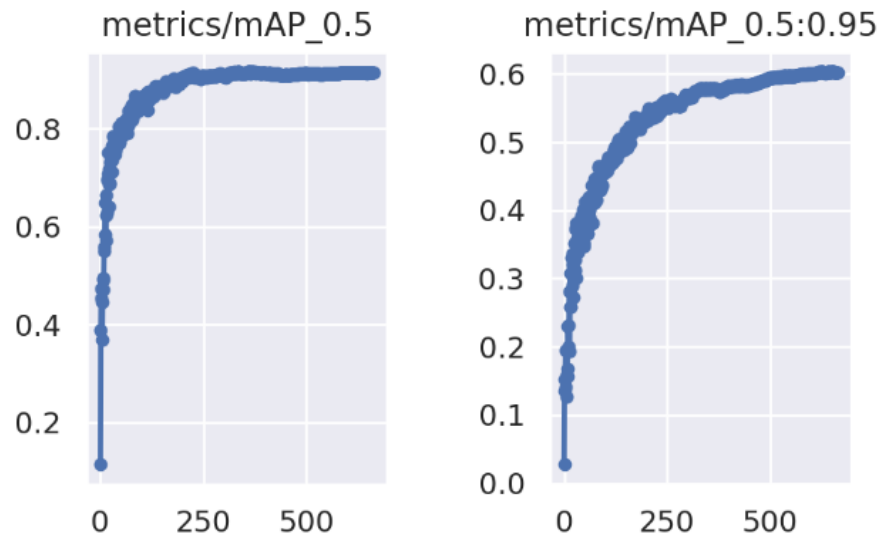


Fig. 3. Illustration of performance metrics for generated YOLOv5-based violence detection model.

V. CONCLUSION AND FUTURE WORKS

A promising area of research that can improve public safety is using deep learning-based approaches for violence detection in IoT-based surveillance systems. This study presented the YOLO algorithm to address the issues related to violence detection. We utilize YOLOv5 as an advanced object identification technique that can quickly and accurately detect multiple objects in an image. We create a model for violence detection using images of violent and non-violent scenes divided into testing, validation, and training sets. The model's performance is evaluated using standard performance indicators. Our system can learn to recognize patterns of violence and accurately differentiate between violent and non-violent classes in real time. Therefore, the proposed method in this study is based on an efficient and fast deep learning architecture named as YOLOv5 network. This network, as previous studies proved and indicated, it is very effective in real-time detection algorithms. Inspiring of this, we also adopted YOLOv5 algorithm and generated a model to deal with violence detection. As our experimental results indicated, the proposed method present accurate results and provide satisfy in real-time requirement. Future work could focus on improving the robustness of the model by addressing various environmental factors that may affect violence detection accuracy. Additionally, the development of larger datasets with diverse scenarios can improve the generalizability of the model. Finally, further investigation could explore the integration of multiple sensors and modalities, such as audio and motion sensors, to enhance the accuracy and reliability of violence detection systems.

REFERENCES

- [1] F. U. M. Ullah, M. S. Obaidat, A. Ullah, K. Muhammad, M. Hijji, and S. W. Baik, "A comprehensive review on vision-based violence detection in surveillance videos," *ACM Comput Surv*, vol. 55, no. 10, pp. 1-44, 2023.
- [2] B. Omarov, S. Narynov, Z. Zhumanov, A. Gumar, and M. Khassanova, "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review," *PeerJ Comput Sci*, vol. 8, p. e920, 2022.
- [3] F. U. M. Ullah *et al.*, "AI-assisted edge vision for violence detection in IoT-based industrial surveillance networks," *IEEE Trans Industr Inform*, vol. 18, no. 8, pp. 5359-5370, 2021.
- [4] M. Islam, A. S. Dukyil, S. Alyahya, and S. Habib, "An IoT Enable Anomaly Detection System for Smart City Surveillance," *Sensors*, vol. 23, no. 4, p. 2358, 2023.
- [5] W. Ullah *et al.*, "Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data," *Future Generation Computer Systems*, vol. 129, pp. 286-297, 2022.
- [6] L. M. Dang, K. Min, H. Wang, M. J. Piran, C. H. Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit*, vol. 108, p. 107561, 2020.
- [7] N. Mumtaz *et al.*, "An overview of violence detection techniques: current challenges and future directions," *Artif Intell Rev*, pp. 1-26, 2022.
- [8] M. H. Rohit, "An IoT based System for Public Transport Surveillance using real-time Data Analysis and Computer Vision," in *2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAECC)*, IEEE, 2020, pp. 1-6.
- [9] M. Ramzan *et al.*, "A review on state-of-the-art violence detection techniques," *IEEE Access*, vol. 7, pp. 107560-107575, 2019.
- [10] A. Singh, T. Anand, S. Sharma, and P. Singh, "IoT based weapons detection system for surveillance and security using YOLOV4," in *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, 2021, pp. 488-493.
- [11] P. Zhou, Q. Ding, H. Luo, and X. Hou, "Violence detection in surveillance video using low-level features," *PLoS One*, vol. 13, no. 10, p. e0203668, 2018.
- [12] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, "Violence detection using oriented violent flows," *Image Vis Comput*, vol. 48, pp. 37-41, 2016.
- [13] M. M. Soliman, M. H. Kamal, M. A. E.-M. Nashed, Y. M. Mostafa, B. S. Chawky, and D. Khattab, "Violence recognition from videos using deep learning techniques," in *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, IEEE, 2019, pp. 80-85.
- [14] P. Wang, P. Wang, and E. Fan, "Violence detection and face recognition based on deep learning," *Pattern Recognit Lett*, vol. 142, pp. 20-24, 2021.
- [15] G. Sreenu and S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *J Big Data*, vol. 6, no. 1, pp. 1-27, 2019.

- [16] S. U. Khan, I. U. Haq, S. Rho, S. W. Baik, and M. Y. Lee, "Cover the violence: A novel Deep-Learning-Based approach towards violence-detection in movies," *Applied Sciences*, vol. 9, no. 22, p. 4963, 2019.
- [17] N. AlDahoul, H. A. Karim, R. Datta, S. Gupta, K. Agrawal, and A. Albunni, "Convolutional Neural Network-Long Short Term Memory based IOT Node for Violence Detection," in *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (ICAIET)*, IEEE, 2021, pp. 1–6.
- [18] D. Choqueluque-Roman and G. Camara-Chavez, "Weakly supervised violence detection in surveillance video," *Sensors*, vol. 22, no. 12, p. 4502, 2022.
- [19] A.-M. R. Abdali and R. F. Al-Tuma, "Robust real-time violence detection in video using cnn and lstm," in *2019 2nd Scientific Conference of Computer Sciences (SCCS)*, IEEE, 2019, pp. 104–108.
- [20] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, p. 217, 2021.

Towards Automated Evaluation of the Quality of Educational Services in HEIs

Silvia Gaftandzhieva¹, Rositsa Doneva², Mariya Zhekova³, George Pashev⁴
University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria^{1,2,4}
University of Food Technology, Plovdiv, Bulgaria³

Abstract—The provision of educational services with high quality is a matter of concern to all stakeholders in higher education (academic staff, administration, students, etc.). According to many researchers, student satisfaction is an indicator of service quality in higher education institutions (HEIs), and evaluating the quality of educational and administrative services from students is an effective tool for improving the quality of HEIs. To ensure a competitive benefit over other educational institutions, HEIs leadership should take measures leading to improved student feedback on the quality of the provided administrative and education services, seek ways to exceed student expectations and provide high-quality services. Due to the great importance of the opinion of students on the quality of the services offered, many HEIs develop and use tools to assess student satisfaction with the quality of the services in the HEI. Little researched in the literature is the issue regarding the need to develop tools for HEIs leadership allowing survey results analysis, tracking trends over the years and comparing HEIs results. Based on a detailed analysis of developed questionnaires for evaluating the quality of services, this paper explores the possibilities of automation of the overall process for conducting questionnaire surveys of student's satisfaction with the quality of services. As a result, a software prototype of a tool to automate the entire process for assessing student satisfaction is proposed - from questionnaire modelling, survey organizing and conducting to the analysis of the collected data. The developed tool allows governing bodies in HEIs to make informed decisions to improve the quality of services and to compare the results with those of competing universities.

Keywords—Quality assurance; higher education; educational services; administrative services; data analysis

I. INTRODUCTION

Quality assurance is a developmental process in the European Higher Education Area (EHEA). Standards, criteria and performance indicators are the starting point in the quality evaluation process at a given time. The implemented quality assurance methodologies of evaluation agencies assess whether or not HEI achieve threshold standards, focus on identifying or promoting HEIs excellence and formulate recommendations for quality improvement. In turn, excellence models set goals for institutions to exceed minimum expectations [1].

Due to the global growth of the higher education sector, HEIs are facing significant challenges to undertake sustainability initiatives in teaching, research and development and administrative services. High competition forces HEIs to review their policies, procedures and marketing guidelines to

ensure that they provide quality educational services and globally recognized education [2-3].

HEIs aim to produce services related to teaching, research and public service [4], divided into two main groups- administrative and educational services. The primary purpose of the administration in the HEI is to enable the performance of the main functions by providing support, integration, coordination, supervision, service of the learning processes, scientific research and public services. Regardless of differences in the specific nature of administrative work, all forms of administrative work can be considered a service. When considering administrative work as a service provision, there are two main groups of users: internal (academic and non-academic staff and students) and external (funding organizations, industry representatives, prospective students and individuals interacting with HEIs). Examples of administrative services offered to students are career guidance services, counselling, participation in internship programs, accommodation in dormitories, and administration of the training conducted [5-8]. On the other hand, teachers provide academic services in a university environment that are directly related to the training provided in the academic disciplines [6]. Gupta and Kaushik [9] note that services have characteristics that distinguish them from products, and assessing their quality can be challenging.

According to Al-Ababneh and Alrhaimi [10], there is no single model of the education quality management system. Even though the effectiveness of higher education management depends on external factors arising in the educational system management, HEIs leadership responds to the educational process effectiveness and the quality of the services provided. It requires the implementation of innovative management methods based on modern information technologies.

The development of technologies in the period of globalization and the industrial revolution 4.0 has had a formidable impact on how organizations from various sectors, including HEIs, perform their daily work [11]. In the digital transformation process, HEIs should invest funds to develop their infrastructure to ensure prestige, meet minimum standards, and use technology to answer the students growing needs. When implementing new solutions, HEI management often must resolve conflicts with existing academic or administrative systems or procedures and staff antipathy to technological innovation. Managing and addressing all these challenges is critical to maintaining the quality of services offered and the effectiveness of HEIs [12]. For this reason, HEIs leadership should integrate all actions for quality

assurance into the management process in HEIs. Improving quality may require changes in various areas, such as human resource management, finance and budget, infrastructure, administrative services, etc. Therefore, to achieve the set goals, it is necessary to consider the assurance and evaluation of the quality of services as tools for the strategic development of HEIs [13]. During the transition from traditional to mixed learning, imposed by the Covid-19 pandemic, problems are emerging in terms of ensuring and maintaining the quality of education, including the compliance of educational programs with the requirements of the labour market, the expectations of students and their families, digital transformation of the educational process and economic stability of institutional networks [14–18].

All stakeholders (academic staff, administration, students, the general public, etc.) are concerned about assuring quality educational services. Quality management in higher education requires understanding the needs of all stakeholders [9], [19–21] and adopting and implementing strategic plans to enable higher education to reply to the needs of external (employers) and internal users (students and teachers) [22-23]. Kazeroony [24] believes many factors explain the need for restructuring strategies to deliver quality education, including the ever-changing characteristics of learners, technological advances and economic reasons. According to Bernhard, due to the demand for quality services, most academic institutions worldwide have undergone a significant transformation [25]. Kettunen believes that the only way to improve stakeholder confidence in the education system is by integrating a plan for quality education services into the overall institutional governance framework and implementing quality assurance systems [26-27].

The service quality in HEIs cannot be examined without considering the student as the primary user of the offered educational and administrative services. Student satisfaction significantly affects the sustainability and development of HEIs, and therefore HEIs leadership should focus on providing better service quality [28-29].

The process of providing quality services is focused on meeting student expectations, continuous improvement and sharing responsibilities [30]. This process ensures HEIs have done well and supported students throughout their training as much as possible. Arokiasamy and Abdullah underlined the need for HEIs to provide a well-rounded university experience due to the potential impact of student satisfaction on HEI competitiveness, student retention, and efforts to attract new students in a highly competitive higher education market [31]. HEI leadership should consider that prospective students gather information about the HEI by contacting current graduates and visiting the website and social media pages. The great satisfied students are more loyal to their HEI and spread positive comments and recommend the HEI to others [32]. Therefore, to ensure that an HEI has a competitive advantage over other HEIs, management should do activities to improve student feedback on the quality of administrative and educational services, seek ways to exceed student expectations and provide high-quality services [33–38]. Students' perspectives on the quality of educational services can be seen as a basis for

adapting marketing efforts to answer student needs [6] and improve institutional performance [39].

According to many researchers, student satisfaction is an indicator of service quality in HEIs, and students' evaluation of the quality of educational services is an effective tool for improving the quality of HEIs [7], [31], [37], [40–52].

Based on a detailed analysis of developed questionnaires for evaluating the quality of services, this paper explores the possibilities of automation of the overall process for studying student satisfaction with the quality of services. Section II reviews different factors contributing to student satisfaction. Section III discusses various methods and tools used to measure student satisfaction. Section IV presents the developed prototype of a software tool for surveys conducted. Section V presents the results, which allows HEIs to make informed decisions to improve the quality of services and compare the results of their HEIs with those of competing universities. Section VI, conclusion, discusses the contributions, limitations of the study and plans for future research.

II. QUALITY OF SERVICES IN HEIS

Kara and DeShields [53] suggest HEIs recognizing the importance of student evaluation of quality services would, in most cases, meet the student's needs to a great extent. Several empirical studies have been conducted over the years on the factors contributing to student satisfaction. Douglas, Douglas and Barnes found [54] the quality of learning resources was not a determining factor. According to the results of other studies, the quality of resources is a vital component of the quality of educational services [55-56]. The quality of resources is a multidimensional construct evaluated by indicators for support teaching, learning and research activities in HEIs. Such indicators include lecture facilities, laboratory facilities, library services and access to information and communication technology (ICT) infrastructure and digital resources [57-58]. According to Arambewela and Hall, teaching style, innovative provision of knowledge, faculty support and feedback influence students' satisfaction [59]. Other researchers also point to innovative changes in curricula [60] and teaching methods [61] as determinants of student satisfaction. Muhsin, Nurkhin, Pramusinto, Afsari and Arham [62] also explored the relationship between university governance, teaching quality and student satisfaction and concluded that teaching quality, teaching facilities and good university governance have a positive and significant impact on student satisfaction. Tuan cites the services provided by the administrative staff, the know-how, skills and attitude of the academic and non-academic staff as leading factors in student satisfaction [63]. Similar are the main aspects influencing the quality of services in HEIs identified by Sultan and Wong [64] - academic environments, academic and non-academic staff. Another group of researchers identified technological, learning, executive and psychological environment as the main factors for student satisfaction [51], [65]. Kara, Tanui and Kalai identified learning facilities, availability of textbooks and library environment as determinants of student satisfaction [66]. Other researchers point to understanding and effective communication between students and staff within the education environment as a leading factor in determining service quality

and student satisfaction [2], [46]. According to Vinogradova, Kulyamina, Vasileva, Bronnikova and Vishnyakov, many factors can influence the formation of consumer expectations, such as the student's own needs, life experience, public opinion, the state of the educational organization, current information on the labour market [67]. According to Hoque, Akhter, Absar, Khandaker and Al-Mamun [29], among the main factors that influence student satisfaction are the comfortable of lecture halls, service timely provision, the capacity of non-academic staff to solve problems, experienced lecturers for teaching and research, and the focus of university management on students.

III. APPROACHES AND TOOLS FOR QUALITY ASSURANCE AND ASSESSMENT

The difficulties in identifying the quality dimensions make challenges in developing models and tools for evaluating the quality of services in higher education [13] and using industry models for service quality evaluation. Due to their limitations, imposed mainly by the centrality of student learning services, industry models can be used with partial success [19] and therefore have to be adapted to reflect the specifics of higher education.

Total quality management (TQM) is a widely used model for improving the performance of service providers and customer satisfaction. Companies that implement TQM [68] transform their organisational culture by engaging all their members to contribute to improving products, processes and services. Many HEIs implement TQM to respond to market pressures while producing high-quality results and striving for self-improvement due to their social contribution [69-70]. Some researchers are sceptical about its applicability in HIEs [71] before the identity and characteristics of education [72-74]. However, many scholars have found that TQM can address stakeholder expectations and challenges in HIEs [75].

The interest in developing models for assessing service quality began in the 1980s. In the beginning, research focused primarily on developing industrial models for evaluating customer satisfaction, such as the perception model customer-perceived service quality [76], SERVQUAL [77] and SERVPERF [78].

Marsh [79-80] proposed the SEEQ tool allowing students to assess the quality of teaching, content and learning in nine areas – Teaching/Course Value, Instructor Enthusiasm, Organization of Presentations and Materials, Group Interaction, Student-Teacher Interaction, Scope, Exams/Assessment, Tasks/Reading Materials, and Workload.

Parasuraman, Zeithaml and Berry [81] proposed a model for measuring service quality based on a multiple-item scale. SERVQUAL is based on the view that the evaluation of the service quality of customers is fundamental, and service quality is closing the gap between service expectations and perceptions. Researchers outline ten dimensions of service quality: Reliability, Responsiveness, Competence, Access, Courtesy, Communication, Trust, Security, Customer understanding and Tangibility. Dimensions in the model are defined as a measure of how well the level of service provided reply customer expectations. To overcome some difficulties in

evaluation, researchers proposed an updated version of the model in 1988. The updated version has five dimensions [77]: Tangible assets (physical facilities, equipment, staff, etc.), Reliability (the ability for reliable service provision), Responsiveness (willingness to serve and assist customers quickly), Security (employees' ability to inspire trust and confidence, politeness and awareness), Empathy (providing individual attention to customers). The evaluation is done through a questionnaire containing 22 questions to assess the five dimensions of the service, allowing for evaluating the customers' expectations and the service provider's performance. Due to its flexibility and ability to be adapted to sector-specific requirements, the proposed model is widely used to assess the quality of services in industries from various sectors, including retail, banking, healthcare and education [82–90]. According to several researchers, the SERVQUAL model is the most well-known and commonly used model for evaluating the quality of services in higher education, incl. from students [9], [39], [91–105].

Tan and Kek [91] used SERVQUAL to assess student satisfaction in Singapore and concluded that some cultural factors should be considered when developing the assessment questionnaires.

Dado, Taborecka-Petrovicova, Riznic and Rajic [92] used SERVQUAL to study the service quality in HEIs in Serbia. They conclude there is a significant gap between student expectations and perceptions. Legcevic, Mujic and Mikrut [94] also used SERVQUAL to identify the gap between students' expectations and perceptions of educational services in Croatia. The survey results show that the negative difference in service dimensions can be used as a guideline for planning and allocating resources to improve the quality of educational services. Over the years, researchers have developed several modified versions of the tool adapted for evaluating the quality of services in HEI.

Aghamolaei and Zare [106] proposed a modified version of the SERVQUAL instrument that allows the evaluation of the quality of educational services by students. The questionnaire measures perceptions and expectations of students from the service in five dimensions – Confidence, Responsiveness, Empathy, Reliability and Tangibility. Study results were analysed using SPSS13 software using descriptive statistics, paired t-test, Wilcoxon, Friedman and ANOVA. Using the proposed tool, 350 students evaluated the service quality.

Zafiroopoulos and Vrana [107] developed a modified version of the model to assess the quality of services in HEIs in Greece from students and teachers. The evaluation study shows no significant differences in how students and academic staff perceive the quality of education.

The HEQUAL tool (based on SERVQUAL) [108] allows students to evaluate the quality of services in HEIs. Students complete a questionnaire with 27 indicators divided into five groups: teaching and course content; administrative services; academic facilities; university infrastructure; support services. The creators underline the possibility of expanding the functionality of the service quality assessment tool by all stakeholders – academic staff, support and administrative staff.

Donlagić and Fazlić [13] developed a service quality assessment tool based on SERVQUAL. The proposed questionnaire contains 25 questions for each scale: one to measure the expectations of students and one to measure their perception of the services provided. The questions cover all dimensions of the SERVQUAL model – tangible assets (four questions for equipment, infrastructure, interior, teaching materials, etc.), reliability (six questions for reliably providing the service, such as allowing student problems, claims and requests), responsiveness (three questions related to the provision of quick service to students), assurance (six questions to evaluate the knowledge and politeness of the academic and non-academic staff and their ability to express trust and confidence) and empathy (six questions for the individual attention given to students). Each question is rated on a 7-point Likert scale. HEIs leadership can use the results of the service quality evaluation as input for planning and strategy setting.

Hassan and Yusof proposed a modified version of SERVQUAL that allows the study of the difference between expectations and satisfaction of students with the quality of educational services [109]. The sub-dimensions of education service quality are Reliability, Assurance, Empathy, Responsiveness, Tangibles (program and service quality), Communication, Knowledge/Expertise, Systems/Secondary Services, Social Responsibility and Development. The authors used a questionnaire to collect the data and a t-test and discriminant technique for results analysis.

Guillén Perales [110] proposed an approach to assess the impact of the primary variable on service quality and determined its significance from students. The evaluation is in two stages. In the first stage, quality evaluation of the service is carried out based on the feedback collected from the students, using a modified version of the SERVQUAL model for this purpose. In the second stage, evaluators should compare the results obtained by students and academic staff. The proposed approach was tested for service quality evaluation by 580 students. The results reveal the most significant dimensions of service quality identified by students. The comparison of the results with those of the academic staff showed notable differences in quality evaluation.

Rizos, Sfakianaki and Kakouris [111] discussed the differences between students' perceptions and expectations of the quality of administrative services. They explored the quality of administrative services of an HEI by assessing student satisfaction in the TQM context. The developed questionnaire follows the SERVQUAL model. It contains 22 questions to research perceptions and expectations regarding the quality of administrative services adapted for the educational environment, divided into five dimensions: Tangibility, Reliability, Responsiveness, Assurance and Empathy. With the proposed tool, the quality evaluation of administrative services of 5 HEIs in Greece was carried out based on primary data from 104 students. The obtained results make it possible to formulate recommendations for importance and effectiveness.

Rozak et al. [18] proposed a model for evaluating the quality of educational services in HEIs based on the SERVQUAL model. Following the proposed model, the

authors developed a questionnaire with 25 closed-ended questions on two scales: one to measure students' expectations regarding the quality of educational services and the other to measure student satisfaction. The collected data is analysed using numerical and statistical analysis techniques and tools such as SPSS. The validity and reliability of each of the items of the model dimensions are measured using the reliability test and the qualification of Cronbach's alpha scores. The collected data should be analysed using descriptive statistics. Using this tool, 236 students evaluated the service quality in Russia and Indonesia.

Hoque, Akhter, Absar, Khandaker, and Al-Mamun [29] developed an instrument to measure student satisfaction with service quality in private universities in Bangladesh based on the SERVQUAL model. The questionnaire developed to collect primary data contains 43 questions – 4 for demographic characteristics, 21 for quality of service, 10 for student satisfaction and 8 for student loyalty to the university. All these questions require a response on a 5-point Likert scale. During the pilot study, 229 students filled in the questionnaire. Primary data were analysed using AMOS 22 and structural equation modelling (SEM).

Ganbold, Park and Hong [112] proposed an approach for evaluating students' requirements regarding the quality of educational services in HEIs based on three models – SERVQUAL, KANO and TIMKO. The evaluation takes place in three phases. During Phase 1, using SERVQUAL, a measurement factor is determined to assess the quality of the educational service. The respondents' perceptions of service quality are classified using a two-dimensional quality classification scheme applying the KANO model. During the last phase, the degree of satisfaction and dissatisfaction of the students is calculated based on the TIMKO equation. This approach provides a satisfactory level of quality indicators to improve student satisfaction based on the PCSI index and ultimately allows HEIs leadership to develop a student satisfaction strategy. The tool was experimented with, to determine the degree of student satisfaction with higher education services in Mongolia and identify the quality characteristics that can improve student satisfaction based on the Potential Customer Satisfaction Improvement Index (PCSI).

The IPA importance-performance analysis is an exciting addition to the existing service quality measurement models [113-114]. According to this model, consumer satisfaction is a function of two components – the importance of the product to the customer and its performance by the service provider. IPA is a diagnostic tool that can identify attribute importance and a product or service benefits to satisfy customer needs [114, 115]. As it diagnoses the main disadvantages and sets the priorities using this tool, companies can overcome the shortcomings of SERVQUAL and discover their strengths and weaknesses. IPA uses a matrix in which one axis measures supplier performance, and the second axis measures customer importance [116]. Due to its simplicity and usefulness in making significant management decisions, IPA is used to evaluate service quality in various fields [117], including higher education [118]. Some researchers fault the model before the applied methods of dividing the quadrants and

evaluating the results. As a result, some modifications have been proposed [113], [119–121].

Researchers doubt the SERVQUAL model because perceptions and expectations are measured together after consumers use the service. On the one hand, this may subconsciously change expectations, and on the other hand, evaluating the service before submitting give often a different result [122-123].

According to Cronin and Taylor, the relationship between expected and received quality is not an appropriate approach for evaluating service quality, and they suggest considering it as a predictor of the service quality only perceptions [78]. The developed SERVPERF model includes 22 items to measure customer satisfaction with service. According to researchers [83], [124] SERVPERF outperforms SERVQUAL in selecting the most effective service quality model in developing countries.

As a result of research, Abdullah [125] researched the general applicability of the SERVPERF model in HEIs. He proposes a modified version for assessing student satisfaction with the services offered. HedPERF includes 49 quality indicators specific to higher education (13 from the SERVPERF model), divided into six dimensions – Non-academic aspects, Access, Academic aspects, Clarity, Reputation and Programmatic issues. Since it is based on SERVPERF, it also assesses service quality as a performance function. The tool was tested for validity and reliability by conducting an empirical study. As disadvantages of the model, researchers point to the overlapping of questions, the emphasis on administrative aspects, its limitations for evaluating other services, and the small number of HEIs in which HedPERF has been tested [88], [126].

Shaik, Lowe and Pinegar [127] proposed a tool to measure the quality of distance learning from students. The developed DL-sQUAL tool allows assessment of the quality of 23 services, divided into three areas – Quality of training services, Management and administrative services and Communication. Based on the evaluation results, administrators can identify services that need to be improved and opportunities for staff training. According to its creators, only administrators of distance learning can use the DL-sQUAL to assess the strengths and weaknesses of the services offered.

Hussain and Birol [6] developed a tool to evaluate student satisfaction with service quality based on SERVQUAL and SEEQ. They suggest three dimensions for quality evaluation: service quality (non-academic services), learning quality (academic services) and student satisfaction. The first two dimensions (non-academic and academic services) are considered multidimensional constructs and independent learning variables, and satisfaction is the dependent learning variable. Service quality is assessed in five areas (tangibility, reliability, responsiveness, confidence and empathy) and learning quality in nine areas (Learning values, Instructor enthusiasm, Course organization, Breadth of coverage, Group interaction, Individual understanding, Examination/assessment rules, Tasks and Workload). The proposed questionnaire contains 59 questions – 22 for service quality (based on SERVQUAL), 33 for learning quality (based on SEEQ), and 4

for student satisfaction. All these questions require a response on a 5-point Likert scale. A pilot study of the tool was conducted in Cyprus involving 330 students. The authors use means, standard deviation and frequencies, reliability analysis, exploratory factor analysis and regression analysis for results analysis.

Adapting the so-called "360-degree feedback" for evaluating the human resources management of a given company has been created as a tool for quality evaluation of management activities in HEIs [128, 129]. The teachers, students and graduates give feedback based on criteria defined according to the evaluated object (curriculum, processes, disciplines, etc.). Each stakeholder evaluates only these criteria for which (s)he has the necessary knowledge or experience.

Kara, Tanui and Kalai [66] explored the relationship between educational quality service and student satisfaction and developed an instrument to evaluate educational quality service and student satisfaction. The questionnaire contains 64 questions – 26 for academic resources, eight for administrative services, 22 for teaching and eight for social services offered. All these questions require a response on a five-point Likert scale. Using the questionnaire, the authors evaluate the quality of services in eight universities in Kenya by collecting primary data from 1062 students. They used factor analysis, descriptive statistics and regression analysis for data analysis.

Based on Harvey and Green's [130] quality framework, Kivistö and Pekkola [4] underlined a possible understanding of the dimensions of quality in HE administration – quality as exclusivity/excellence, quality as perfection/consistency, quality as fitness for purpose, quality as value for money, quality as transformation. According to them, the main tools for ensuring the quality of administrative services are regulations and action plans, administration audits, conducting periodic surveys among the users of administrative services (academic staff, non-academic staff, students, external stakeholders), analysis of quantitative data for financial and human resources and cost measurement, performing benchmarking and conducting internal forums for open dialogue and sharing of experience on the use of administrative services.

Vnoučková, Urbancová and Smolová [131] evaluated key internal quality management processes from students and identified factors for effective internal quality process management. They offered a tool for the quality of the management process evaluation in five key areas – leadership and strategic planning, focus on students and stakeholders, measurement of student learning outcomes, human resources planning and education process management. The authors used a quantitative study (filling in questionnaires) and a qualitative study within the target groups to collect data for evaluation. Students rated all indicators in the questionnaire on a five-point Likert scale. Primary data from the questionnaires are analysed using descriptive statistics and bivariate statistical methods.

According to Lestari and Khusaini [46], analytical tools can support HEIs in fulfilling their vision and mission. They are suitable for measuring student satisfaction and can be used to evaluate the quality of educational services. They offer a tool for assessing student satisfaction with the quality of academic

and non-academic administrative services and the availability of educational facilities. The evaluation is going on a proposed model with indicators in 5 areas – Reliability, Responsiveness, Confidence, Empathy and Physical evidence. The primary data for conducting the study were collected by filling in questionnaires from 184 students who evaluated the indicators using a five-point Likert scale. The survey results are analysed by comparing the differences in expectations and satisfaction in using the service, conducting a matched pairs test using SPSS and showing in a Cartesian diagram.

Prima and Saputra [132] considered the level of satisfaction of customers as a measure of the quality of services in HEIs. They propose a service quality assessment model based on previous research in the field [133] with ten dimensions (reality, responsiveness, competence, access, courtesy, communication, reliability, security, customer understanding/knowledge and tangibility) divided into five main areas – Reliability, Responsiveness, Assurance (competence, courtesy, security), Empathy (access, communication and understanding of the customer) and Tangibility.

Mastoi, Xin Hai and Saengkrod [3] explored the level of student satisfaction with the quality of administrative services, educational services, support facilities and physical environment. Based on an extensive literature review and qualitative data collection from interviews conducted with students and faculty, they identified five main dimensions of HESQUAL that were considered independent determinants for evaluating a dependent variable for overall student satisfaction – administrative quality, physical environment quality, the primary educational quality, the quality of the support facilities and the transformative quality. They collected the data for conducting the study from 500 questionnaires, did results analyses in SPSS and used multiple linear regression analysis to evaluate the role played by each factor in predicting student satisfaction.

Amoako and Asamoah-Gyimah [51] explored the factors contributing to student satisfaction with educational services and the quality assurance of services offered by HEIs. They offered two instruments – to evaluate the overall satisfaction of teachers and students from the quality of services. The questionnaire for students was developed based on previous research by Stukalina [65] and contains 19 questions divided into three dimensions - Technological environment (assesses the availability, adequacy and access to modern technologies in the context of their studies), Learning Environment (assesses the situation in the classroom and the teaching approach), The psychological environment (evaluates belonging to the academic family). The developed instrument assessed the satisfaction with educational services of 1500 students in Ghana. Researchers used Analysis of Moment Structures (AMOS) to validate the tool and test the hypotheses.

Montemayor [134] studied the ongoing procedures, prevailing practices and beliefs, conditions for existing relationships, perceived effects, and developmental trends. This process goes beyond simple data collection and tabulation. Primary data for the study were collected using a questionnaire and survey results were processed with SPSS v. 23.

Lian and Putra [11] proposed a methodology for evaluating the effectiveness and role of educational administration in HEIs in the digital era. They suggested a quantitative approach to measure data for efficacy and a qualitative approach to analyse the data according to the role. For the quantitative analysis, a questionnaire was developed to evaluate the administration with questions in four areas – goal achievement (effectiveness of the set goals), system (availability of resources and the connection with the external environment), strategic groups (level of satisfaction) and competitive values (criteria for success with educational administrative factors such as educational facilities, infrastructure, finance and environment). Each question requires a response on a five-point Likert scale. Qualitative research is conducted through observations, literature studies and interviews. The proposed approach has been used to evaluate the effectiveness of educational administration at PGRI Palembang University, Indonesia.

Vinogradova, Kulyamina, Vasileva, Bronnikova and Vishnyakov [67] identified criteria and indicators for evaluating educational services and developed a methodology for measuring the quality of educational services in HEIs. They proposed 33 quantitative indicators for quality evaluation, divided into five areas – Educational programs (10), Teaching staff (6), Educational technologies (6), Material and technical provision of the educational process (5), and Management of education processes (6). They define weights and formulas for calculating the score for each area and indicator. Based on the indicators' scores, they calculate a composite factor of the quality of educational services as considered the area weight in the calculation formula). In this way, the composite coefficient makes it possible to evaluate the quality of educational services in quantitative terms, the maximum value of which is 1. The proposed methodology allows objective evaluation and helps the HEIs leadership to take measures to improve the quality of educational services.

Krymets, Saienko, Bilyakovska, Zakharov, and Ivanova [23] proposed an approach to determining the requirements for the quality of education from the perspective of administrative staff, students and employers, developed based on stakeholder theory and TQM. The approach involves a survey with sets of questions for different stakeholder groups – Administrative/support staff (14 items), Teaching staff (19 items), Students (26 items), and Industry (15 items). They developed four frameworks with requirements to meet the needs of all users of educational services and to ensure the evaluation of the overall quality of higher education. Each question requires a response on a five-point Likert scale. For each statement, employer respondents rated both the expectations of graduates and the actual student performance in the workplace. The survey results were processed with Statistica 22.0 using basic analysis methods – Cronbach's α to check the reliability of the constructed sets of questions and Pearson's correlation to assess the reliability of perception and stakeholder requirements analysis.

Tran [135] offered a tool to assess students' perception of the quality of educational services with 22 questions divided into five areas – Educational services (four for admission services, transfer, fees, etc.), Facilities and equipment (four for classrooms, equipment, teaching aids, level of safety and

hygiene), Educational environment (five for attitude, enthusiasm and correctness of teachers during educational activities), Educational activities (four for training activities), Development and progress of students (five for evaluating learning results). All questions require a response on a five-point Likert scale. During the pilot evaluation, the authors used SPSS software for results analysis.

Hai [136] investigated the factors influencing student satisfaction with service quality in HEIs. To conduct the study, Hai collected data from 396 students. During structured discussions, participants are presented with a list of factors and asked to give their opinion on the listed factors and add some missing factors. Hai used SPSS 20, Cronbach's Alpha reliability coefficient, EFA, CFA and SEM for results analysis.

The results show that six factors influence student satisfaction with the quality of services – teaching staff, facilities, serviceability, educational activities, student support activities and educational programs.

Assiri [137] explored the most significant technical, human, economic, social and administrative obstacles and requirements to make suggestions for using e-government to improve the quality of education services in Saudi Arabia. He developed a tool to identify difficulties in implementing e-administration in HEIs from the perspective of employees and teachers.

Table I summarizes the criteria and the evaluation target (Expectation/Satisfaction) of the studied models for quality evaluation. The comparison proves that many factors affect the quality of the services offered in the education system.

TABLE I. COMPARISON OF STUDIED MODELS

Approach Authors	Criteria	Expectations/Satisfaction
Students Evaluations of Educational Quality (SEEQ) Marsh 1982, 1987	Training/Course Value Enthusiasm of Instructors Organization of Presentations and Materials Group Interaction Student-Teacher Relationship; Exams/Assessments, Assignments/Reading Materials, Workload.	Satisfaction
SERVQUAL Parasuraman, Zeithaml и Berry 1985, 1988	Reliability, Responsiveness, Competence, Access, Courtesy, Communication, Trust, Security, Customer Understanding and Tangibility.	Expectations and satisfaction
Modified SERVQUAL models - Tan, Kek 2004; Dado et al., 2011; Legcevic et al., 2012; Aghamolaei, Zare,2008; Zafiroopoulos & Vrana, 2008	Confidence, Responsiveness, Empathy, Reliability and Tangibility.	Expectations and satisfaction
HEdQUAL Icli & Anil, 2014	Teaching and Course content; Administrative services; Academic facilities; University infrastructure and support services.	Expectations and satisfaction
Donlagić & Fazlić, 2015	Tangible assets (equipment, infrastructure, interior, teaching materials, etc.) reliability (reliable service delivery, resolution of student problems, claims and requests), responsiveness (quick service), confidence (knowledge and courtesy of academic and non-academic staff, expression of trust and confidence) and empathy (given individual attention).	Expectations and satisfaction
Hassan & Yusof 2015	Reliability, confidence, empathy, responsiveness, tangibles (program and service quality), communication, knowledge (expertise), systems (secondary services), social responsibility and development.	Expectations and satisfaction
Rizos et al., 2022	Tangibility, reliability, responsiveness, confidence and empathy.	Expectations and satisfaction
Rozak et al., 2022	Confidence, Responsiveness, Empathy, Reliability and Tangibility	Expectations and satisfaction
Hoque et al., 2023	Demographic characteristics, Quality of service, Student satisfaction and Student loyalty to the university.	Satisfaction
Importance performance analysis IPA Abalo 2007; Sever 2015	Importance of an item to the customer, Benefits of a product or service to meet customer needs, and Performance by the service provider.	Satisfaction
SERVPERF Cronin & Taylor, 1992	22 items to measure customer satisfaction.	Satisfaction
HedPERF Abdullah, 2006	Non-academic aspects, access, academic aspects, clear understanding, reputation and programmatic issues.	Satisfaction
DL-sQUAL Shaik et al., 2006	Quality of training services, Management and administrative services and Communication.	Expectations and satisfaction
Hussain & Birol, 2011	Quality of services (non-academic), Quality of teaching (academic), Student satisfaction.	Satisfaction
Kara, Tanui & Kalai 2016	Quality of academic resources, Quality of administrative services, Teaching and of the social services offered.	Satisfaction
Vnoučková et al., 2018	Leadership and strategic planning, Student and stakeholder focus, Measurement of student learning outcomes, Human resource planning and management of the educational process.	Satisfaction
Lestari & Khusaini, 2018	Reliability, responsiveness, confidence, empathy and physical (material) evidence.	Expectations and satisfaction
Prima & Saputra, 2019	Reliability, Responsiveness, Assurance (competence, courtesy, reliability and security), empathy (access, communication and understanding of the customer) and Tangibility.	Satisfaction
HESQUAL Mastoi et al., 2019	Administrative quality, Physical environment quality, Basic educational quality, Facilities quality, Transformative quality.	Satisfaction
Amoako et al., 2020	Technological Environment, Learning Environment, Psychological Environment	Satisfaction
Vinogradova et al., 2021	Educational programs (10 indicators), Teaching staff (6 indicators), Educational technologies (6 indicators), Material and technical provision of the educational process (5 indicators), and Management of educational processes (6 indicators).	Satisfaction

Krymets et al., 2021	Administrative (support) staff (14 items grouped into 4 factors), teaching staff (19 items into 5 factors), Students (26 items into 5 factors), and Industry (15 items into 4 factors).	Expectations and satisfaction
Tran et al., 2022	Educational services (4 indicators), Facilities and equipment (4 indicators), Educational environment (5 indicators), Educational activities (4 indicators), and Development and progress of students (5 indicators).	Satisfaction
Hai, 2022	Faculty, Facilities, Service capacity, Educational activities, Student support, and Educational programs.	Satisfaction
Assiri, 2023	Technical, Human, Economic, Social and administrative obstacles and requirements for using e-government.	Satisfaction

Several test evaluations of the quality of services in specific HEIs have been carried out using the developed tools. Part of these surveys was organized using software solutions for surveys, such as Google Forms. The conduction of similar surveys with such tools has some disadvantages – the possibility of providing access to the survey questionnaire to external persons, manual data processing when detailed analysis of results is necessary, difficulties in tracking trends in assessments, etc.

Few studies have addressed the issue of monitoring the results of conducted studies [138-139]. None of the considered tools automates the overall process of evaluating student satisfaction, comparing results of individual HEIs and generating recommendations that can support HEIs leadership in decision-making. To ensure the high quality of the services, it is vital HEIs leadership not only to conduct periodic surveys, the results of which should be made public, but also to implement tools that analyse results and present them in a summarized form and allow them to make informed decisions to improve the quality of the services offered. Based on the results, HEIs leaders can identify weaknesses and take measures to improve problem areas to answer the needs of students and ensure student satisfaction with the quality of service provided by the institution's employees.

Despite the various factors, all models have a two-level hierarchical structure and require evaluation on a defined scale (in most cases five- or seven-point Likert scale). This fact enables HEIs leadership to search for solutions to automate the whole process of evaluating the quality of services in HEIs, from conducting surveys, and survey results analysis to the generation of evaluation reports.

IV. SOFTWARE TOOL PROTOTYPE

Automating the overall process for evaluating the quality of educational services from students requires the design, development and implementation of a software tool that allows:

- modelling of a questionnaire for evaluating the quality of educational services, including assigning weights to evaluated indicators;
- provide an opportunity for students to fill in questionnaires;
- generation of reports with the survey results for a specific HEI;
- generation of recommendations for improving the quality of educational services offered in HEIs;
- generation of reports for comparing the results of different HEIs.

The project for a software tool for evaluating student satisfaction with the quality of services offered includes the following six subsystems:

- Subsystem 1: Conceptual modelling of questionnaires (areas, indicators, weights) for quality evaluation;
- Subsystem 2: Modelling and managing quality evaluation procedures in specific HEIs;
- Subsystem 3: Evaluation of the quality of services by students according to the modelled questionnaire;
- Subsystem 4: Modelling of report templates for summarizing the evaluation results;
- Subsystem 5: Generation of reports (for individual HEIs and summary reports) for evaluating the quality of services in HEIs;
- Subsystem 6: Generating recommendations for improving the quality of services.

The developed software prototype UQCS is an online tool for evaluating the quality of educational services in HEIs by students. The tool generates recommendations and reports with evaluation results, allowing HEIs leadership to make informed decisions for improving the quality of services.

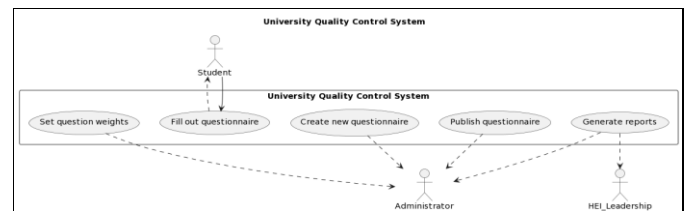


Fig. 1. The UML diagram of the university quality control system (UQCS).

The UML diagram shows (see Fig. 1) the main actors in the system (Student, Administrator and HEIs leadership) and the use cases they can perform. The use cases are:

- UC1: Create new questionnaire;
- UC2: Set question weights;
- UC3: Publish questionnaire;
- UC4: Fill out questionnaire;
- UC5: Generate reports.

Administrators can create a new questionnaire (UC1) by specifying the questions that will be included in the questionnaire and setting the weights for all questions (UC2) that determine how much each question contributes to the overall score of the questionnaire. Once the Administrators create the questionnaire and set the weights of the questions, they can publish the questionnaire so that students can start

filling it out (UC3). The Student can complete the questionnaire by answering questions and submitting their answers (UC4). Administrators and HEIs leadership can generate reports summarizing the evaluation results and recommendations (UC5). They can use these reports to track the quality of education at the university and identify areas where HEIs leadership can make improvements. The management of HEIs can review the evaluation results and the generated recommendations and compare the achievements of the HEI they manage with those of other universities. The arrows between the actors and use cases show the relationships between them. For example, the arrow from Student to UC4 shows that the Student can perform the Fill out questionnaire use case.

Subsystem 1 allows the administrator to create questionnaires that evaluate the quality of services provided in HEIs and assign weights to each question to reflect its importance. Using this subsystem, the administrator can model all the questionnaires considered in Section III. These are just a few of the screens that users see in Subsystem 1.

- *Questionnaire preview screen*: Users see this screen to preview a questionnaire before it is published. They can see how the questionnaire will look and how the questions will be presented;
- *Questionnaire publishing screen*: Users see this screen to publish a questionnaire. Once a questionnaire is published, it will be available for students to fill in.

Insert University

Name:

Location:

Country:

Website:

Phone:

Fig. 2. Insert form for university data.

Subsystem 2 allows administrators to create and manage quality assessment procedures for specific HEIs. The performed procedures can include different stages, such as distributing questionnaires, collecting responses, processing responses, and generating reports. The subsystem is sending emails to HEIs leadership about the organized survey. Users see the *Questionnaire creation screen* when they want to create a new questionnaire. They can enter the title of the questionnaire, add questions, and preview the questionnaire before publishing it. Fig. 2 shows the screen for insert of data for a new University in the Database.

Subsystem 3 allows students to fill in questionnaires to assess the quality of services provided in the HEI. The questionnaires are distributed within the quality assessment procedures created in Subsystem 2. After completing the questionnaire, the student's answers are recorded in the system database. Each student can fill in the questionnaire only once.

The database UCQS contains the following tables:

- *Questionnaire Table* – stores information about the questionnaires created by the administrators;
- *Question Table* – stores the questions associated with each questionnaire;
- *Student Response Table* – stores the responses of students to each question;
- *University Table* – stores information about different HEIs.

Fig. 3 the most significant part of the Python code using Flask that provides an HTTP API for the main functions of the "University Quality Control" system. This API allows users to create questionnaires, submit student responses, and retrieve results and recommendations.

```
from flask import Flask, request, jsonify
app = Flask(__name__)
# Sample data storage (you should use a database in a real application)
questionnaires = []
responses = []
# Endpoint to create a new questionnaire
@app.route('/api/questionnaires', methods=['POST'])
def create_questionnaire():
    data = request.get_json()
    title = data.get('title')
    description = data.get('description')
    questions = data.get('questions')
    if not title or not description or not questions:
        return jsonify({'message': 'Incomplete data. Title, description, and
questions are required.'}), 400
    questionnaire = {
        'title': title,
        'description': description,
        'questions': questions
    }
    questionnaires.append(questionnaire)
    return jsonify({'message': 'Questionnaire created successfully.'}), 201
# Endpoint to submit student responses
@app.route('/api/responses', methods=['POST'])
def submit_response():
    data = request.get_json()
    questionnaire_id = data.get('questionnaire_id')
    student_id = data.get('student_id')
    responses = data.get('responses')
    if not questionnaire_id or not student_id or not responses:
        return jsonify({'message': 'Incomplete data. Questionnaire ID, student ID,
and responses are required.'}), 400
    response = {
        'questionnaire_id': questionnaire_id,
        'student_id': student_id,
        'responses': responses
    }
    responses.append(response)
    return jsonify({'message': 'Student response submitted successfully.'}), 201
# Endpoint to retrieve questionnaire results
@app.route('/api/questionnaires/<int:questionnaire_id>/results',
methods=['GET'])
def get_questionnaire_results(questionnaire_id):
    questionnaire = next((q for q in questionnaires if q['questionnaire_id'] ==
questionnaire_id), None)
    if not questionnaire:
        return jsonify({'message': 'Questionnaire not found.'}), 404
    # Calculate results based on responses (you'll need to implement this logic)
    results = calculate_questionnaire_results(questionnaire_id)
    return jsonify(results), 200
# Endpoint to generate recommendations
```

```
@app.route('/api/questionnaires/<int:questionnaire_id>/recommendations',
methods=['GET'])
def get_recommendations(questionnaire_id):
    questionnaire = next((q for q in questionnaires if q['questionnaire_id'] ==
questionnaire_id), None)
    if not questionnaire:
        return jsonify({'message': 'Questionnaire not found.'}), 404
    # Generate recommendations based on results (you'll need to implement
this logic)
    recommendations = generate_recommendations(questionnaire_id)
    return jsonify(recommendations), 200
def calculate_questionnaire_results(questionnaire_id):
    return {'questionnaire_id': questionnaire_id, 'results': {'question1': 4,
'question2': 3}}
def generate_recommendations(questionnaire_id):
    return {'questionnaire_id': questionnaire_id, 'recommendations':
['Improve teaching methods', 'Enhance student support']}
if __name__ == '__main__':
    app.run(debug=True)
```

Fig. 3. Part of software code.

In the beginning, the necessary modules are imported, including Flask, and an instance of the Flask application is created. Lists in memory (questionnaires and answers) are used for temporary data storage. In a real-world application, you need to use a database to store the data. Here we define four API endpoints using the @app.route() decorator: /api/questionnaires: POST endpoint to create a new questionnaire; /api/responses: POST endpoint to submit student responses; /api/questionnaires/<questionnaire_id>/results: GET endpoint to retrieve questionnaire results; /api/questionnaires/<questionnaire_id>/recommendations: GET endpoint to generate recommendations. The create_questionnaire() endpoint allows the administrator to create a new questionnaire by providing a title, description, and a list of questions. The submit_response() endpoint allows students to submit their responses to a specific questionnaire by providing the questionnaire ID, student ID, and responses. The get_questionnaire_results() endpoint retrieves the results for a specific questionnaire, calculated based on the submitted responses. The get_recommendations() endpoint generates recommendations for improving the quality of services based on the results of a specific questionnaire. The calculate_questionnaire_results() and generate_recommendations() are sample functions representing the logic for calculating questionnaire results and generating recommendations. You should replace them with the actual implementation based on your requirements. The if __name__ == '__main__': block runs the Flask application in debug mode.

Fig. 4 is a part of the Python Flask code that verifies the student user using a simple username and password combination. In a real-world application, one would typically use a more secure authentication mechanism such as JWT (JSON Web Tokens) or OAuth2. We import the necessary modules, including Flask, and create an instance of the Flask app. We use an in-memory dictionary (student_credentials) to store the student usernames and passwords. In a real-world application, you should use a database and securely hash the passwords. We define a route /login using the @app.route() decorator. This route expects a POST request with a JSON payload containing the username and password. The login() function handles the login request. It checks if the provided username and password match the credentials stored in

student_credentials. If the login is successful, the function returns a JSON response with a success message and an authentication token. In this example, we are using a simple string as the token, but in a real application, you should use JWT or a similar authentication token mechanism. If the login fails (incorrect username or password), the function returns a JSON response with an error message.

```
from flask import Flask, request, jsonify
app = Flask(__name__)
# Sample student credentials (replace with actual credentials or use a
database)
student_credentials = {
    #...
}
# Sample authentication token (replace with JWT or OAuth2 token in a real
application)
def generate_token(username):
    return f'TOKEN_{username}'
# Route to handle student login
@app.route('/login', methods=['POST'])
def login():
    data = request.get_json()
    username = data.get('username')
    password = data.get('password')
    if not username or not password:
        return jsonify({'message': 'Username and password are required.'}), 400
    # Check if the provided username and password match the stored credentials
if username in student_credentials and student_credentials[username] ==
password:
        token = generate_token(username)
        return jsonify({'message': 'Login successful.', 'token': token}), 200
    else:
        return jsonify({'message': 'Invalid username or password.'}), 401
```

Fig. 4. Part of software code for verification.

Subsystem 4 allows users to create report templates that summarize the results of evaluating the quality of services. We used Jasper Reports Server as a reporting tool to implement Subsystem 4, which involves modelling report templates with evaluation results. Using JasperSoft Studio, we designed four report templates and defined their corresponding parameters (see Table II).

TABLE II. A LIST OF DEVELOPED TEMPLATES OF REPORTS

Template	Parameter	Visualized data
Detail results of HEI	Survey ID HEI ID Survey period	Average scores by evaluated indicators (questions)
Summarized results of HEI	Survey ID HEI ID	Average scores by evaluated areas
HEIs ranking	Survey ID Survey period	Calculated average grades of HEIs
Detail HEIs ranking	Survey ID Survey period	Calculated average grades of HEIs for each evaluated area

During this stage, we designed SQL queries and data adapters to retrieve evaluation results from the UQCS database and populate the elements of report templates. After the user input parameter values, JasperSoft Studio fills in all data storage elements of templates with data retrieved from the UQCS database stored from Subsystem 3. The calculation of average scores is embedded in the developed document templates. The formula used considers both the grades given by the students on each indicator (question) and the assigned weights.

Subsystem 5 allows users to generate reports to assess the quality of services provided in HEIs. They can generate reports for an individual HEI or a group of HEIs. Subsystem 5 can run all developed templates (developed within Subsystem 4) stored in the Jasper Report Server. For this to be possible, a connection is required between Subsystem 3 (that store the evaluation results) and the Jasper Report Server. Based on this integration, the subsystem passes data to Jasper Reports Server for report generation. To enable report generation from the UQCS, a mechanism to trigger the report generation based on user requests has been implemented in Subsystem 5. Before report generation, the user must select the name of the report template and submit parameter values. The value of the HEI ID parameter is passed by the UQCS tool to eliminate the possibility of generating a report with the results of another HEI. The user must input values for other parameters (Survey ID and Survey period) as select sequentially values from drop-down lists. After receiving parameter values, JasperReport Server retrieves data from the UQCS database, calculates the results and fills in the report template with data. Then, JasperReport Server returns a completed report to the UQCS tool. The UQCS display reports on the screen and allows users to download them in the desired format (e.g., HTML, DOCX, XLSX, PDF, CSV) and share it with different stakeholders. Fig. 5 presents a part of the Python Flask client code that interacts with the Jasper Reports Server to generate and download a report.

```
import requests
app = Flask(__name__)
# Function to generate and download a report from Jasper Reports Server
def generate_report(report_template, parameters):
    jasper_server_url = 'http://jasper_reports_server_url'
    username = 'your_jasper_username'
    password = 'your_jasper_password'
    # Authenticate with Jasper Reports Server
    auth_url = f'{jasper_server_url}/jasperserver/rest_v2/login'
    auth_data = {'j_username': username, 'j_password': password}
    auth_response = requests.post(auth_url, data=auth_data)
    if auth_response.status_code != 200:
        return 'Authentication Failed.', 401
    # Generate the report
    report_url =
f'{jasper_server_url}/jasperserver/rest_v2/reports/{report_template}'
    headers = {'Authorization': f'Basic {auth_response.text}',
              'Content-Type': 'application/json'}
    report_response = requests.post(report_url, headers=headers,
    json=parameters)
    if report_response.status_code != 200:
        return 'Report Generation Failed.', 500
    # Download the report
    download_url = report_response.json()['outputResource']['uri']
    download_response = requests.get(download_url, headers=headers)
    if download_response.status_code == 200:
        # Save the report to a local file
        with open('generated_report.pdf', 'wb') as file:
            file.write(download_response.content)
        return 'Report Generated Successfully.', 200
    else:
        return 'Report Download Failed.', 500
# .... Some code omitted
```

Fig. 5. Part of software code for interaction with Jasper Reports Server.

The Flask client code provides an endpoint /generate_report that triggers the generation and download of a report from the Jasper Reports Server. The generate_report() function handles

the interaction with Jasper Reports Server. It performs authentication using the provided username and password and generates the report using the specified report template and parameters. The trigger_report_generation() route demonstrates how to trigger the report generation. Replace your_report_template_name with the actual name of the report template on Jasper Reports Server, and value1 and value2 with the required parameters. The generated report is saved locally as generated_report.pdf.

Subsystem 6 allows HEIs leadership and the administrator to generate recommendations for improving the quality of services provided. The recommendations are based on the results of the quality assessment. The Subsystem selects all evaluated areas with a result score of less than four and generates a recommendation for it. The recommendation is generated using Google Bard Artificial Language Model and the Python bardapi Library (see Fig. 6). This subsystem makes use of the following screens:

- *Questionnaire results screen:* Users see this screen to view the results of a questionnaire. They can see how students responded to the questions and the overall score of the questionnaire.
- *Recommendations screen:* Users see this screen to view recommendations for improving the quality of services based on the results of a questionnaire.

```
import bardapi
def generate_recommendations(areas, scores):
    """
    Generates recommendations for improving the quality of services in a
    university based on the evaluated areas and scores.
    Args:
        areas: A list of areas.
        scores: A list of scores for each area.
    Returns:
        A list of recommendations.
    """
    recommendations = []
    for i in range(len(areas)):
        if scores[i] < 4:
            recommendation = "Improve " + areas[i]
            explanation = bardapi.generate_explanation(areas[i])
            recommendations.append((recommendation, explanation))
    return recommendations
```

Fig. 6. Some of the code of the recommendations generator.

This would return the following example list of recommendations (see Fig. 7):

```
[("Improve instructional quality", "The instructional quality can be improved by hiring more qualified professors, providing more resources for students, and creating a more supportive learning environment."), ("Improve student-faculty interaction", "The student-faculty interaction can be improved by creating more opportunities for students to interact with professors, providing more support for student-led initiatives, and creating a more welcoming and inclusive environment."), ("Improve curriculum", "The curriculum can be improved by making sure that the courses are relevant to the needs of students, providing more opportunities for hands-on learning, and ensuring that the curriculum is aligned with the university's mission.")]
```

Fig. 7. Example list of recommendations for instructional quality improved.

V. RESULTS

The software tool UQCS was tested to assess the quality of services in three universities. After completing questionnaires created using Subsystem 1, users generated some reports with the evaluation results.

Here are some screenshots of reports generated by a user with the role “HEIs leadership” during the pilot testing of the tool. Data for experimenting were collected from completed questionnaires for evaluating the quality of services in nine areas (Instructional Quality, Student-faculty Interaction, Curriculum, Support Services, Campus Environment, Value for money, Quality of life, Student diversity, Career opportunities) by students from three universities. The Likert scale values for each question are on a scale of 1 to 5, with 1 being strongly disagree and 5 being strongly agree. The total score for each university is calculated by adding up the Likert scale values for all 10 evaluated areas.

Fig. 8 presents the generated report with summary results of one HEI who participated in the experiment. It shows the calculated average marks for each evaluated area and the overall student satisfaction mark. Based on the results, the HEI leadership can gain insights into which areas the university shows poor results and make informed decisions for improving the quality of services in these areas.

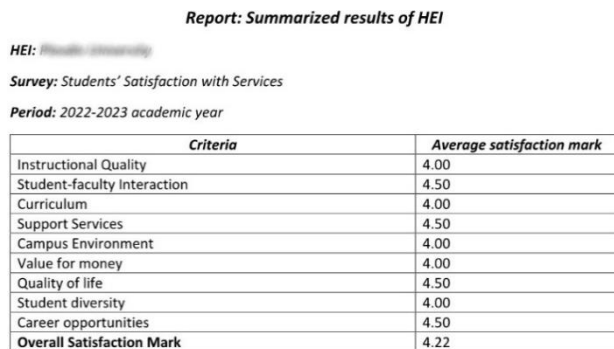


Fig. 8. Summarized results of HEI.

Fig. 9 shows a generated report with calculated overall satisfaction marks of all HEIs who participated in the experiment. The calculated scores allow the results of HEIs to be compared and their leaders to make informed decisions to improve the quality of services, which will lead to a rise in the ranking and an increase in the prestige of the HEI.

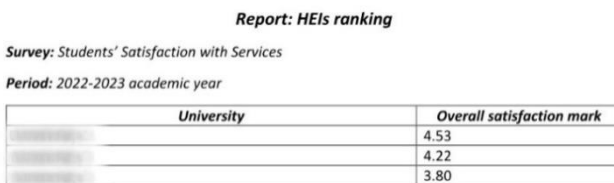


Fig. 9. HEIs ranking.

Fig. 10 shows a generated report with recommendations for improving the quality of services in one of the evaluated universities.

Recommendations for Improving Educational Service Quality

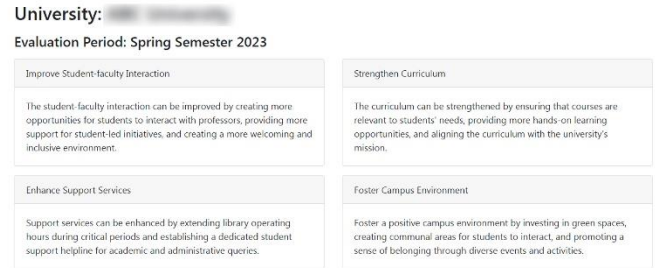


Fig. 10. Generated recommendations.

VI. CONCLUSION

The proposed approach automates the overall process for measuring student satisfaction with the quality of educational and administrative services in HEIs. The developed tool automatically analyzes the collected data. HEIs leadership can use it to generate summary reports with survey results allowing them to track the satisfaction of their students and compare the results with those of competing higher education institutions. The summarized results, and the recommendations generated by the tool, allow managers to make informed decisions to improve the quality of services. The results of the experimental testing of the developed prototype of the software tool prove its applicability to support the HEIs leadership in making decisions for improving the quality of the offered educational services.

The conducted research also has some limitations. Since it has been tested with students from a small number of universities, it does not allow making general conclusions about the overall student satisfaction with the quality of educational and administrative services in higher education institutions.

In the future, the tool's functionalities will be expanded by:

- enriching the set of report templates with the results of conducted studies, including for comparative analysis across multiple HEIs, enabling institutions to benchmark their performance against others;
- extending the report generation capabilities to allow users to customize report templates, select specific data points, and choose visualizations for more tailored insights;
- improving the user interface and experience of the UQCS tool to make it more intuitive and user-friendly for both administrators and HEIs leadership;
- implementing data validation and cleaning mechanisms to ensure that the input data for evaluations is accurate and consistent, leading to more reliable results;
- integrating machine learning models to predict potential areas of improvement based on historical data and trends;
- strengthening the security aspects of the system, including encryption of sensitive data, role-based access

control, and secure communication with external servers;

- optimizing the software architecture to ensure scalability as more HEIs adopt the tool and the user base grows;
- implementing a feedback mechanism within the tool to collect user suggestions and experiences, driving continuous enhancements.

ACKNOWLEDGMENT

This paper is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01. The paper reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

- [1] M. Brusoni, R. Damian, J. Sauri, S. Jackson, H. Kömürçügil, M. Malmedy & L. Zobel, The concept of excellence in higher education, 2006.
- [2] H. Uiso, J. Magali, "Service quality variables for assessing students' satisfaction in higher learning institutions: evidence from empirical literature reviews," *The Pan-African Journal of Business Management*, 1(1): 2017, pp. 71-82
- [3] A. Mastoi, L. XinHai & W. Saengkrod, "Higher education service quality based on students' satisfaction in Pakistan," *European Scientific Journal ESJ*, 75(11), 2019, pp. 32-62.
- [4] J. Kivistö & E. Pekkola, Quality of administration in higher education, 2017.
- [5] R. Emanuel & J. Adams, "Assessing college student perceptions of instructor customer service via the quality of instructor service to students (QISS) questionnaire," *Assessment & Evaluation in Higher Education*, 31(5), 2006, pp. 535-549.
- [6] K. Hussain & C. Birol, "The assessment of non-academic and academic service quality in higher education," *Egitim Arastirmalari-Eurasian Journal of Educational Research*, 42, 2011, pp. 95-116.
- [7] S. Khoo & S. McGregor, "Service quality and student/customer satisfaction in the private tertiary education sector in Singapore," *International Journal of Educational Management*, 31(4), 2017, pp. 430-444.
- [8] I. Mawudeku & F. J. Amkumah, "The Role of Administrators in Enhancing Students' Academic Performance in Higher Education Institutions," *IOSR Journal of Humanities and Social Science (IOSR-JHSS)* Volume, 26, 2021, pp. 16-22.
- [9] P. Gupta & N. Kaushik, "Dimensions of service quality in higher education – critical review (students' perspective)," *International Journal of Educational Management*, 32(4), 2018, pp. 580-605.
- [10] H.A. Al-Ababneh & S.A. Alrhaimi, "Modern approaches to education management to ensure the quality of educational services". *TEM Journal*, 9(2), 770, 2020.
- [11] B. Lian and M. J. Putra, "The Role and Effectiveness of Education Administration in Developing Institutions of Higher Education in the Digital Era," 4th International Conference on Education and Social Science Research (ICESRE), *KnE Social Sciences*, 2022, pp. 592–602.
- [12] M. Fakhruhnizam, R. Abdullah, M. A. Jabar, R. Nor Haizan, Nor Aida Abdul Rahman, "Towards The Integration of Quality Management System and Knowledge Management System In Higher Education Institution: Development Of Q-Edge Kms Model," *Acta Information Malaysia*, 2(2), 2018, pp. 04-09.
- [13] S. Đonlagić & S. Fazlić, "Quality assessment in higher education using the SERVQUALQ model," *Management: journal of contemporary management issues*, 20(1), 2015, pp. 39-57.
- [14] S. V. Veretkhina, V. A. Novikova, "Transformation of education in the digital economy," *Contemporary Problems of Social Work*. Vol 5. № 2 (18), 2019, pp. 30-37.
- [15] A. Zhang, G. Li, Z. Li, "Research on High Quality Development of Higher Education," 5th International Conference on Humanities and Social Science Research (ICHSSR) Guilin, China. 319, 2019, pp. 472-475.
- [16] J. L. Carcamo, "Factors associated with the teacher's practice that affect the dropout of students in the e-learning modality, a case study in the context of Chilean higher education," *CIENCIA E INGENIERIA*. 42 (1), 2021, pp. 101-112.
- [17] J.L. Gonzalez-Geraldo, F. Monroy, B. Del Rincon Igea, "Impact of a Spanish higher education teacher development programme on approaches to teaching," *Psychometric properties of the s-ati-20 scale. Educacion XXI*. 24 (1), 2021, pp. 213-232.
- [18] L. A. Rozak, M. Bahri Arifin, I.N. Rykova, O.A. Grishina, A. Komariah, D. Nurdin & O.V. Dudnik, "Empirical evaluation of educational service quality in the current higher education system," *Emerging Science Journal*, 6, 2022, pp. 55-77.
- [19] N. Becket & M. Brookes, "Quality management practice in higher education - what quality are we actually enhancing?" *Journal of Hospitality Leisure Sport and Tourism Education*, 7(1), 2008, pp. 40-54.
- [20] C.R. Santos, A.M. Contreras, C. Faúndez & G.F. Palomo-Vélez, "Adapting the SERVQUAL model to a physical activity break satisfaction scale," *International Journal of Workplace Health Management*, 8(1), 2015, pp. 34-45.
- [21] S. Sahney, "Use of multiple methodologies for developing a customer-oriented model of total quality management in higher education," *International Journal of Educational Management*, 30(3), 2016, pp. 3260-353.
- [22] A. Moosavi, M. Mohseni, H. Ziaifar, S. Azami-Aghdash, M.G. Manshadi & A. Rezapour, "The quality of educational services from students' viewpoint in Iran: a systematic review and meta-analysis," *Iranian Journal of public health*, 46(4), 2017, page 13.
- [23] L.V. Krymets, O.H. Saienko, O.O. Bilyakovska, O.Y. Zakharov & D.H. Ivanova, "Methodology for Ensuring Quality Management of Higher Education," *Revista geintec-gestao inovacao e tecnologias*, 11(3), 2021, pp. 945-959.
- [24] H. Kazeroony, *The strategic management of higher education: Serving students as customers for institutional growth*, USA: Business Expert Press, 2012.
- [25] A. Bernhard, *Quality assurance in an international higher education area: A case study approach and comparative analysis*. Germany: Springer Fachmedien, 2012.
- [26] J. Kettunen, J., "A conceptual framework to help evaluate the quality of institutional performance," *Quality Assurance in Education* 16(4), 2008, pp. 322–332.
- [27] J. Kettunen, "Strategy and quality maps in higher education," *US-China Education Review* 8(2), 2011, pp. 149–156.
- [28] I. Alnawas, "Student orientation in higher education: development of the construct," *Higher Education*, 69, 2015, pp. 625-652.
- [29] U.S. Hoque, N. Akhter, N. Absar, M.U. Khandaker & A. Al-Mamun, "Assessing Service Quality Using SERVQUAL Model: An Empirical Study on Some Private Universities in Bangladesh," *Trends in Higher Education*, 2(1), 2023, pp. 255-269.
- [30] M. Alifuddin, *Reformasi Pendidikan: Strategi Informatif Peningkatan Mutu Pendidikan*, Magna Script. Jakarta, 2012.
- [31] A. Arokiasamy & A. Abdullah, "Service quality and students' satisfaction at higher learning institutions: A case study of Malaysian university competitiveness," *International Journal of Management and Strategy*, 3(5), 2012, pp. 1-16.
- [32] H. Alves, M. Raposo, "The measurement of the construct satisfaction in higher education," *Serv. Ind. J.* 2009, 29, pp. 203–218.
- [33] P. Stevens, "Dineserv: A tool for measuring service quality in restaurants," *Cornell Hotel. Restaur. Adm. Q.* 1995, 36, pp. 56–60.
- [34] I.M. Tahir, "Importance-performance analysis of service quality among business students: An exploratory study," *Interdiscip. J. Contemp. Res. Bus.* 2010, 2, pp. 330–341.

- [35] I. G. Lee, *The Effect of Perception of Educational Service Quality on Re-Enrollment Intention and Word-of-Mouth Intention of College Students*, Ph.D. Thesis, Busan, Republic of Korea, 2012.
- [36] S. Annamdevula, R.S. Bellamkonda, "Effect of student perceived service quality on student satisfaction, loyalty and motivation in Indian university," *J. Serv. Manag.* 2016, 11, pp. 488–517.
- [37] A. Burgess, C. Senior & E. Moores, "A 10-year case study on the changing determinants of University student satisfaction in the UK," *Public Library of Sciences (PLoS ONE)*, 13(2), 2018, pp. 1-15.
- [38] R. Togholi, F. Moradi, L. Hassani, T. Aghamolaei, N. Mehedi, F. Mahmoodi & B. Ziapour, "Evaluation of the educational services quality from the viewpoint of postgraduate students at Kermanshah University of medical sciences in 2019," *Journal of Education and Health Promotion*, 10. 2021.
- [39] M. Khanli, H. Daneshmandi & A. Choobineh, "The students' viewpoint on the quality gap in educational services," *Journal of advances in medical education & professionalism*, 2(3), 114, 2014.
- [40] F.M. Hill, "Managing service quality in higher education: the role of the student as primary consumer," *Quality Assurance in Education*, 3 (3), 1995, pp. 10-21.
- [41] R. Barnett, "The marketized university: Defending the indefensible. In *The Marketisation of Higher Education and the Student as Consumer*," Molesworth, Routledge: Oxfordshire, UK, 2011; pp. 39–52.
- [42] J.I. Yadav, "Service quality towards student satisfaction: An empirical investigation in selected management institutions of Sangli City," *Ninth AIMS International Conference on Management*, 2012.
- [43] V. Okogbaa, "Quality in higher education: the need for feedback from students," *Journal of Education and Practice*, 7 (32), 2016, pp. 139-143.
- [44] A. Rouf, M. Rahman and M. Uddin, "Students' satisfaction and service quality of HEIs," *International Journal of Academic Research in Business and Social Sciences*, 6(5), 2016, pp. 376-390.
- [45] A. Azam, "Service quality dimensions and students' satisfaction: a study of Saudi Arabian private higher education institutions," *European online journal of Natural and social sciences*, 7(2), 2018, pp. 275-284.
- [46] S. Lestari & K. Khusaini, *Analysis of Student Satisfaction on Academic and Non Academic Administration*, 2018.
- [47] D. Napatipulu, R. Rahim, D. Abdullah, M.I. Setiawan, S. Abdillah, A.S. Ahmar, J. Simarmata, R. Hidayat, H. Nurdianto and A. Pranolo, "Analysis of student satisfaction toward quality of service facility," *Journal of Physics: Conference. Series*, 954(1), 2018, pp. 1-7.
- [48] G. Prakash, "Quality in higher education institutions: insights from the literature," *The TQM Journal*, 23(3), 2018, pp. 250-261.
- [49] J.L. Gregory, "Applying SERVQUAL: Using service quality perceptions to improve student satisfaction and program image," *Journal of Applied Research in Higher Education*, 11(4), 2019, pp. 788-799.
- [50] H.L. Gao, "Understanding the impact of administrative service quality on satisfaction and loyalty towards university students." *Higher Education Research*, 5(1), 2020, pp. 25-30.
- [51] I. Amoako & K. Asamoah-Gyimah, "Indicators of students' satisfaction of quality education services in some selected universities in Ghana," *South African Journal of Higher Education*, 34(5), 2020, pp. 61-72.
- [52] P. Kaur & K. Amanpreet, "Service quality in higher education: a literature review," *Elementary Education Online*, 19(40), 2020, pp. 6308-6324.
- [53] A. Kara and O. DeShields. *Business student satisfaction, intentions and retention in higher education: An empirical investigation*. http://www.marketingpower.com/Community/ARC/gated/Documents/Teaching/MEO/student_satisfaction.pdf, 2004.
- [54] J. Douglas, A. Douglas, B. Barnes. "Measuring student satisfaction at a UK university," *Quality Assurance in Education* 14(3), 2006, pp. 251–267.
- [55] H. Encabo, "Canonical correlation analysis of student perception on instructional quality and satisfaction," *JPAIR Multidisciplinary Journal* 6, 2011, pp. 1–16.
- [56] R. Prasad & M. Jha, "Quality measures in higher education: A review and conceptual model," *Quest Journals Journal of Research in Business and Management*, 1(3), 2013, pp. 23 – 40.
- [57] C. Taib, A. Warokka & H. Hilman, "The library's quality management system and quality assurance in higher education: A lesson from Southeast emerging educational hub," *Communications of the IBIMA*, 2012, pp. 1-11.
- [58] V. Mahmood, M. Dangi & K. Ali, "Exploring students' contentment level of the infrastructure at a public higher education institution in Malaysia," *Gading Business and Management Journal*, 18(1), 2014, pp. 61-82.
- [59] R. Arambewela and J. Hall, "An empirical model of international student satisfaction," *Asian Pacific Journal of Marketing and Logistics* 21(4), 2009, pp. 555–569.
- [60] E. Razinkina, I. Pankova, E. Trostinskaya, L. Pozdeeva Evseeva and A. Tanova, "Student satisfaction as an element of education quality monitoring in innovative higher education institution," *E3S Web of Conferences*, 33, 03043. 2018.
- [61] A. Osman and R. Saputra, "A pragmatic model of student satisfaction: A viewpoint of private higher education," *Quality Assurance in Education* 27(2), 2019, pp. 142–165.
- [62] S. Muhsin, A. Nurkhin, H. Pramusinto, N. Afsari and A. F. Arham. "The relationship of good university governance and student satisfaction," *International Journal of Higher Education* 9(1), 2020.
- [63] N. Tuan, "Effects of service quality and price fairness on student satisfaction," *International Journal of Business and Social Science* 3(19), 2012, pp. 132–150.
- [64] P. Sultan and H.Y. Wong, "Antecedents and consequences of service quality in a higher education context: A qualitative research approach," *Quality Assurance in Education*, 21(1), 2013, pp. 70-95.
- [65] Y. Stukalina, "Identifying predictors of student satisfaction and student motivation in the framework of assuring quality in the delivery of higher education services," *Business, Management and Education* 12(1), 2014, pp. 127–137.
- [66] A. Kara, E. Tanui & J. Kalai, *Quality of academic resources and students' satisfaction in public universities in Kenya*, 2016.
- [67] M. Vinogradova, O. Kulyamina, L. Vasileva, E. Bronnikova & V. Vishnyakova, "Methodology for assessing the quality of educational services in higher education," *Revista on line de Política e Gestão Educacional*, 2021.
- [68] W.E. Deming, *Out of the crisis*. MIT Centre for Advanced Engineering Study, 1986.
- [69] A. Weckenmann, G. Akkasoglu & T. Werner, "Quality management – history and Trends," *The TQM Journal*, 27(3), 2015, pp. 281–293.
- [70] E. Psomas & J. Antony, "Total quality management elements and results in higher education institutions: The Greek case," *Quality Assurance in Education*, 25(2), 2018, pp. 206-223.
- [71] F. Cruz, I. Gálvez & R. Santaolalla, "Impact of quality management systems on teaching-learning processes," *Quality Assurance in Education*, 24(3), 2016, pp. 394-415.
- [72] N. Bouranta, E. Psomas & J. Antony, "Findings of quality management studies in primary and secondary education: A systematic literature review," *The TQM Journal*, 33(3), 2021, pp. 729-769.
- [73] E. Sfakianaki, "A measurement instrument for implementing total quality management in Greek primary and secondary education," *International Journal of Educational Management*, 33(5), 2019, pp. 1065-1081.
- [74] V.M. Sunder, "Constructs of quality in higher education services," *International Journal of Productivity and Performance Management*, 65(8), 2016, pp. 1091-1111.
- [75] N. Mehta, P. Verma & N. Seth, "Total quality management implementation in engineering education in India: An interpretive structural modelling approach," *Total Quality Management & Business Excellence*, 25(1-2), 2014, pp. 124-140.
- [76] C. Grönroos, "A service quality model and its marketing implications. *European Journal of Marketing*", Vol. 18, No. 4, 1984, pp. 36-44.
- [77] A. Parasuraman, V.A. Zeithaml & L.L. Berry, "SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, 64(1), 1988, pp. 12-40.
- [78] J.J. Cronin & S.A. Taylor, "SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions minus-expectations

- measurement of service quality,” *Journal of Marketing*, 58(1), 1992, pp. 125-131.
- [79] H.W. Marsh, “SEEQ: A reliable, valid and useful instrument for collecting students’ evaluations of university teaching,” *British Journal of Educational Psychology*, 52(1), 1982, pp. 77-95.
- [80] H.W. Marsh, “Students’ evaluations of university teaching: Research findings, methodological issues, and directions for further research,” *Journal of Educational Research*, 11(3), 1987, pp. 253-388.
- [81] A. Parasuraman, V.A. Zeithaml & L.L. Berry, “A conceptual model of service quality and its implication. *Journal of Marketing*, 49, 1985, 41-50.
- [82] R. Narang, “How do management students perceive the quality of education in public institutions?” *Quality Assurance in Education*, 20(4), 2012.
- [83] M.A. Adil, O.F. Ghaswyneh and A.M. Albkour “SERVQUAL and SERVPERF: a review of measures in services marketing research,” *Global Journal of Management and Business Research Marketing*, 13(6), 2013, pp. 65-76.
- [84] H.M. Awoke, “Service quality and consumer satisfaction: Empirical evidence from saving account consumers of the banking industry,” *European Journal of Business and Management*, 7(1), 2015, pp. 144-164.
- [85] K. Randheer, “Service quality performance scale in higher education: culture as a new dimension,” *International Business Research*, 8(3), 2015, pp. 29-41.
- [86] R. Galeeva, “SERVQUAL application and adaptation for educational service quality assessments in Russian higher education,” *Quality Assurance in Education*, 24(3), 2016, pp. 329-348.
- [87] P. Jain and V.S. Aggarwal, “Service quality models: a review,” *BVIMSR’s Journal of Management Research*, 7 (2), 2015, pp. 125-136.
- [88] E.O. Onditi and T.W. Wechuli, “Service quality and student satisfaction in higher education institutions: a review of the literature,” *International Journal of Scientific and Research Publications*, 7(7), 2017, pp. 328-335.
- [89] F. Khattab “Developing a service quality model for private higher education institutions in Lebanon,” *Journal of Management and Marketing Review*, 3(1), 2018, pp. 24-33.
- [90] N. Ramya, A. Kowsalya and K. Dharanipriya, “Service quality and its dimensions,” *EPRA International Journal of Research and Development (IJRD)*, 4(2), 2019, pp. 38-41.
- [91] K.C. Tan, S.W. Kek, “Service quality in Higher education using and enhanced SERVQUAL approach,” *Quality in Higher education*, 10 (1), 2004, pp. 17-24.
- [92] J. Dado, J. Taborecka-Petrovicova, D. Riznic & T. Rajic, “An empirical investigation into the construct of higher education service quality,” *International Review of Management and Marketing*, 1(3), 2011, pp. 30-42.
- [93] M. Javadi, *Quality Assessment for Academic Services in University of Isfahan According to the Student’s Opinion Using SERVQUAL. Model Interdisciplinary Journal of Contemporary Research in Business. Vol. 3 Edisi 4*, 2011.
- [94] J. Legcevic, N. Mujic & M. Mikrut, “Kvalimetar mjerni instrument za upravljanje kvalitetom na Sveučilištu u Osijeku,” *International scientific symposium “quality and social responsibility”*. In *Croatian Association of Quality Managers*, 2012, pp. 271–283.
- [95] C.E. Nell and M.C. Cant, “SERVQUAL: Student’s perception and satisfaction regarding the quality of service provided by student administration departments within tertiary institutions,” *Corporate Ownership & Control*, 11(4), 2014, pp. 242-249.
- [96] P. Green, “Measuring service quality in higher education: A South African case study,” *Journal of International Education Research*, 10(2), 2014, pp. 131 - 142.
- [97] S. Anwowie, J. Amoako and A. Abrefa, “Assessment of students’ satisfaction of service quality in Takoradi Polytechnic: the students’ perspective,” *Journal of Education and Practice*, 6 (29), 2015, pp. 148-155.
- [98] V. Teeroovengadum, T.J. Kamalanabhan & A.K. Seebaluck, “Measuring service quality in higher education: Development of a hierarchical model (HESQUAL),” *Quality Assurance in Education*, 24(2), 2016, pp. 244-258.
- [99] K.F. Latif, I. Latif, U.F. Sahibzada, M. Ullah, “In search of quality: measuring higher education service quality (HiEduQual),” *Total Quality Management & Business Excellence*, 2017, pp. 1 – 24.
- [100] A. Nsamba and M. Makoe, “Evaluating quality of students’ support services in open distance learning,” *Journal of Distance Education*, 18(4), 2017, pp. 91 – 103.
- [101] M. Saleem, A. Hussain and S. Ahmad, “Identification of gaps in service quality in higher education,” *Bulletin of Education and Research*, 39 (2), 2017, pp. 171 – 182.
- [102] A.S. Williams, *An exploratory study of students’ expectations and perceptions of service quality in a South African higher education institution*. M.Tech., Rhodes University, 2018.
- [103] R. Rezaee, Z. Yazdani, Z., Zahedani & N. Zarifsanaiy, “Quality of educational services: students’ point of view,” *Journal of Advanced Pharmacy Education & Research*, 8(S2), 163, 2018.
- [104] L. Neyra, E. Espinoza & A. Ramírez, “Quality of educational service at the Faculty of Social Sciences and Humanities of a Public University,” *Educação & Formação*, 6(3), 2021.
- [105] S. Daryazadeh, M. Yavari, M. Sharif, M. Azadchahr, S. Hoseini & H. Akbari, “Assessing the Quality of Educational Services from the Viewpoint of Clinical Teachers and Medical Students Using SERVQUAL Model,” *Educational Research in Medical Sciences*, 11(2), 2022.
- [106] T. Aghamolaei, S. Zare, “Quality gap of educational services in viewpoints of students in Hormozgan University of medical sciences,” *BMC medical education*, 8(1), 2008, pp. 1-6.
- [107] C. Zafiroopoulos & V. Vrana, “Service Quality Assessment in a Greek Higher Education Institute,” *Journal of Business Economics and Management*, 9 (1), 2008, pp. 33-45.
- [108] G. Icli & N. Anil, “The HEDQUAL scale: A new measurement scale of service quality for MBA programs in higher education,” *South African Journal of Business Management*, 45(3), 2014, pp. 31-43.
- [109] Z. Hassan, A. Raheem, M. Yusof, “Educational Service Quality at Public Higher Educational Institutions: Difference between Perceived Service and Expected Service,” *Journal of Economics, Business and Management*, Vol. 3, No. 11, 2015.
- [110] A. Guillén Perales, F. Liébana-Cabanillas, J. Sánchez-Fernández & L.J. Herrera, “Assessing university students’ perception of academic quality using machine learning,” *Applied Computing and Informatics*. 2020.
- [111] S. Rizos, E. Sfakianaki & A. Kakouris, “Quality of administrative services in higher education,” *European Journal of Educational Management*, 5(2), 2022, pp. 115-128.
- [112] B. Ganbold, K. Park, J. Hong, *Study of Educational Service Quality in Mongolian Universities. Sustainability* 2023, 15, 580, 2023.
- [113] J. Abalo, J. Varela & V. Manzano, “Importance values for Importance–Performance Analysis: A formula for spreading out values derived from preference rankings,” *Journal of Business Research*, 60(2), 2007, pp. 115-121.
- [114] I. Sever, “Importance-performance analysis: A valid management tool?” *Tourism Management*, 48, 2015, pp. 43-53.
- [115] F. Pai, T. Yeh & C. Tang, “Classifying restaurant service quality attributes by using Kano model and IPA approach,” *Total Quality Management & Business Excellence*, 29(3-4), 2016, pp. 301–328.
- [116] R. Dabestani, A. Shahin, M. Saljoughian & H. Shirouyehzad, “Importance performance analysis of service quality dimensions for the customer groups segmented by DEA,” *International Journal of Quality & Reliability Management*, 33(2), 2016, pp. 160-177.
- [117] N. Slack, N. “The importance-performance matrix as a determinant of improvement priority,” *International Journal of Operations & Production Management*, 14(5), 1994, pp. 59-75.
- [118] M. Joseph & B. Joseph, “Service quality in education: a student perspective. *Quality Assurance in Education*, 5(1), 1997, pp. 15–21.
- [119] K. Matzler, F. Bailom, H. Hinterhuber, B. Renzl & J. Pichler, “The asymmetric relationship between attribute-level performance and overall customer satisfaction: a reconsideration of the importance–performance analysis,” *Industrial marketing management*, 33(4), 2004, pp. 271-277.

- [120]M. Feng, J. Mangan, C. Wong, M. Xu & C. Lalwani, "Investigating the different approaches to importance–performance analysis," *The Service Industries Journal*, 34(12), 2014, pp. 1021-1041.
- [121]S. Ormanovic, A. Ciric, M. Talovic, H. Alic, J. Eldin & D. Causevic, "ImportancePerformance Analysis: Different Approaches," *Acta Kinesiologica*, 11(2), 2017.
- [122]J.M. Carman, "Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions. *Journal of Retailing*, 66(1), 1990, 33-55.
- [123]C. Grönroos, "Toward a third phase in service quality research," *Advances in Services Marketing and Management: Research and Practice*, 2, 1993, pp. 49-64.
- [124]R. Bolton & J. Drew, "A multi-stage model of customers' assessments of service quality and value," *Journal of Consumer Research*, 17, 1991, pp. 375-84.
- [125]F. Abdullah, "Measuring service quality in higher education: HEDPERF versus SERVPERF." *Marketing Intelligence and Planning*, 24(1), 2006. pp. 31-47.
- [126]K. Brunson, "Examining the Need for Customized Satisfaction Survey Instruments for Measuring Brand Loyalty for Higher Educational Institutions," *Liberty University School of Business Journal*, 2010, 1-22.
- [127]N. Shaik, S. Lowe, K. Pinegar, "DL-sQUAL: A multiple-item scale for measuring service quality of online distance learning programs", *Online Journal of Distance Learning Administration*, IX(II), 2006.
- [128]M. Calatrava Moreno, *Towards a Flexible Assessment of Higher Education with 360-degree Feedback, Information Technology Based Higher Education and Training (ITHET)*, 2013.
- [129]D. Tee, P. Ahmed, "360 degree feedback: an integrative framework for learning and assessment," *Teaching in Higher Education*, Vol. 19, Issue 6, 2014.
- [130]L. Harvey & D. Green, "Defining quality. Assessment and Evaluation in Higher Education," 18(1), 1993, pp. 9–34.
- [131]L. Vnoučková, H. Urbancová, H. Smolová, "Internal quality process management evaluation in higher education by students", *DANUBE: Law, Economics and Social Issues Review*, ISSN 1804-8285, De Gruyter, Warsaw, Vol. 9, Iss. 2, 2018, pp. 63-80.
- [132]W. Prima & R. Saputra, "Designing an information system model of academic service based on customer relationship management at university," In *Journal of Physics: Conference Series* (Vol. 1387, No. 1, p. 012009). IOP Publishing, 2019.
- [133]V. Zeithaml, M. Bitner, and D. Gremler, *Services Marketing: Integrating Customer Focus Across the Firm*. McGraw-Hill Education, 2017.
- [134]C.T. Montemayor, "Innovation and Quality Management in Higher Education," *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(9), 2020, pp. 1559-1579.
- [135]H. Tran, M. Nguyen, T. Nguyen, H. Tran, "Students Satisfaction with Public Services in Higher Education Institutions: The Case of Vietnam," *Specialusis Ugdymas*, 1(43), 2022, pp. 1021-1046.
- [136]N.C. Hai, "Factors Affecting Student Satisfaction with Higher Education Service Quality in Vietnam," *European Journal of Educational Research*, 11(1), 2022, pp. 339-351.
- [137]F. Assiri, *Utilizing E-administration to Improve the Quality of Educational Services at Saudi Universities During COVID-19*, 2023.
- [138]M. Lubis, M. Hasibuan & R. Andreswari, "Satisfaction Measurement in the Blended Learning System of the University: The Literacy Mediated-Discourses (LM-D) Framework," *Sustainability*, 14(19), 12929, 2022.
- [139]L. Grebennikov & M. Shah, "Monitoring trends in student satisfaction," *Tertiary Education and Management*, 19, 2013, pp. 301-322.

Machine-Learning-based User Behavior Classification for Improving Security Awareness Provision

Alaa Al-Mashhour, Dr.Areej Alhogail

Department of Information Systems-College of Computer and Information Science, King Saud University Riyadh, Saudi Arabia

Abstract—Users of information technology are regarded as essential components of information security. Users' lack of cybersecurity awareness can result in external and internal security attacks and threats in any organization that has several users or employees. Although various security methods have been designed to protect organizations from external intrusions and attacks, the human factor is also essential because security risks by "insiders" can occur due to a lack of awareness. Therefore, instead of general nontargeted security training, comprehensive cybersecurity awareness should be provided based on employees' online behavior. This study seeks to provide a machine-learning-based model that provides user behavior analysis in which organizations can profile their employees by analyzing their online behavior to classify them into different classes and, thus, help provide them with appropriate awareness sessions and training. The model proposed in this paper will be evaluated and assessed through its implementation on a sample dataset that reflects users' online activities over a specific period to measure the model's accuracy and effectiveness. A comparison between six classification techniques has been made, and random forest classification had the best performance regarding classification accuracy and performance time. After users are classified, each group can be provided with the appropriate training material. This study will stimulate additional research in this area, which has not been widely investigated, and it will provide a useful point of reference for other studies. Additionally, it should provide insightful information to help decision-makers in organizations provide necessary and effective security awareness.

Keywords—Machine learning; user behavior analysis; cybersecurity; classification; security awareness

I. INTRODUCTION

The internet plays a significant role in many aspects of our lives, and many daily tasks have been digitalized and are required to be completed online. Besides this, the number of users and employees with varying levels of security knowledge and different backgrounds who are required to work online has increased, which has, in turn, influenced organizations' security requirements. Because of this, every organization now has internal cybersecurity and data and asset safety as a priority. Organizations that handle sensitive information assets can operate effectively, locally, and globally, exchanging information quickly and seamlessly among their employees, partners, suppliers, and customers. Indeed, many organizations now rely on online information exchange to keep their operations running smoothly in

collaboration with other parties. However, confidential information is becoming increasingly vulnerable to internal and external security attacks [1]. Although hardware and software-based technologies have been implemented, such as firewalls, proxy servers, and antivirus software, these solutions have not significantly reduced security attacks.

Security attacks or breaches, when they are carried out successfully in organizations, affect inside assets or data. However, the consequences are frequently financial and reputational, undermining customer trust. Applying technical control and systems in this regard is essential. Still, technical controls are only the first line of defense in cybersecurity, and they cannot prevent insiders with elevated access from violating security policies. Many previous studies in this field have discussed the human factor in cybersecurity and the significant role that employees can play in information security breaches. This has increased organizational focus on human threats [2,3].

As a result, many organizations have started to provide cybersecurity awareness training to their employees to make them conscious of cybersecurity threats or any other related issues. Awareness sessions and training are critical to ensuring that staff members act responsibly and are aware of the potential consequences of their online behavior [4]. Due to the importance of cybersecurity awareness inside an organization, various studies have reported that they can become considerably more secure against both internal and external security threats with improved security awareness programs [5–7]. Ryu et al. [8] outlined that a strong awareness program is essential to guarantee that employees properly comprehend their respective internet technology (IT) security duties and roles to safeguard the IT resources delegated to them. Therefore, to reach this level of awareness and responsibility in this regard, awareness sessions on cybersecurity's importance are vital to ensuring the enhancement of the security culture within an organization.

Many employers provide cybersecurity awareness sessions and frequently send out relevant material and emails, as will be viewed in Section II. Nevertheless, these conventional methods are ineffective because tailored and targeted security awareness materials based on the needs and knowledge of employees is required as the level of awareness varies greatly among employees.

This study proposes a machine-learning-based model that enables organizations to analyze users' online behavior,

activities, and actions to target them with appropriate security awareness materials. First, we will investigate six machine learning (ML) classification models to select the most appropriate classifier based on the performance measurements and how accurate each classifier is in forming each user class. We will go through several phases to train and test the models. Additionally, for added validation, we will conduct a cross-validation test to ensure accurate results.

Furthermore, based on the comparison results obtained through the performance calculation using confusion matrix, accuracy, F1, and other measures, including performance time, the best classification technique will be used to classify users into three classes and subsequently target them with suitable awareness sessions. The user classes are the malicious, suspicious, and normal (which require the fewest targeted awareness sessions) behavior classes.

Users' online behavior can reveal much about their knowledge level about cybersecurity and what type of security threats they may cause for their organization, as well as what type of security awareness training must be provided to them. Therefore, we used a dataset, which will be discussed later, that consists of web links that users have visited to show their web-behavior. After user classification, the organization can choose the suitable cyber security awareness materials and session content for each user's class, and it will be saved for subsequent users in the backend database to be sent again by the machine to each particular class without any human interaction.

The proposed model can be implemented as a plug-in for the security operations center dashboard. Therefore, in addition to having the ability to monitor network traffic, endpoints, logs, and security events, the organization will also be able to classify its employees into specific classes to send them classified training materials and take the required action in this regard. In addition, these classes can benefit decision-makers in assessing the organization's weaknesses regarding employees' behavior to define new awareness strategies, IT usage policies, and, if required, new tasks and responsibilities.

The remainder of this paper is organized as follows: The literature review will take place in Section II. Section III comprises the proposed methodology. Section IV presents the result, then a discussion and comparison of the results achieved in Section V, followed by the conclusion Section VI, which concludes the proposed model and presents directions for future work.

II. LITERATURE REVIEW

The use of Internet technology (IT) has increased dramatically since its advent. The rapid increase in internet traffic has led researchers to consider the significance of cybersecurity, and research on the values and methods of cybersecurity awareness has attracted substantial attention. Nevertheless, only a few studies have been conducted on the use of machine learning in cybersecurity awareness. This section covers background knowledge and related work regarding the proposed method.

As we are looking to enhance the user awareness level due to its importance, In fact, traditional training methods, such

as classroom discussions and exercises, have demonstrated their efficacy in increasing trainee awareness and, consequently, their ability to detect issues such as phishing or hacking attempts [9]. However, due to the high cost and number of trainees, traditional class sessions are rendered insufficient and cannot provide the information that individual employees need. Bernaschina et al. [10] studied some security training sessions that concentrated on phishing emails. At the end of each session, the trainees were given a survey to complete the evaluation of usefulness of the previous session as a learning opportunity. These trainees reported that they already had prior knowledge of phishing emails, which demonstrates that nontargeted sessions that are not based on specific behavior lead to the wastage of both time and money, as well as a reduction in benefits for the organization and its employees. Therefore, targeted sessions based on behavior analysis must be created.

Crume et al. [9] found that targeted employee awareness programs based on web behavior can aid in preventing the misuse of an organization's assets. Furthermore, implementing this training will result in numerous benefits for organizations, including improved resource utilization, employee knowledge and performance, and organizational policies and procedures.

Current research on awareness has tended to focus on analyzing users' behaviors based on qualitative data collected through interviews, scales, questionnaires, and surveys. User behavior analysis related to cybersecurity awareness, however, focuses on analyzing users' activities, such as accessing websites and files and user identity. User behavior analysis has successfully identified usage patterns that may indicate unusual or anomalous internet behavior.

A study carried out by Gartner has been mentioned in the work of Kumar and Singh [12] that defined user behavior analysis as outlining and incongruity recognition, which depends on a variety of analytic methodologies, typically combining fundamental analytical methods. Examples of this are policies that influence signatures, pattern recognition, mapping, basic rules of statistics, and advanced analytics tools. However, these methods do not provide accurate data regarding users' real online behavior.

As shown in some of the previous research on the impact of online behavior, this emphasizes the need for organizations to target their employees with specific awareness sessions based on an analysis of those behaviors. Targeted security awareness refers to the provision of training based on the threat that some employees' online behavior may pose. These employees can be identified using behavioral analysis of each user within an organization, using a range of qualitative and quantitative data. Multiple scales are used to assess employee awareness. For example, a Portuguese healthcare institution case study assessed employees' professional awareness of information security by assessing their attitudes and behavior related to cybersecurity [13]. The study consisted of applying and validating scales, such as the risky cybersecurity behaviors (RScB) scale, which is a questionnaire for employees that evaluates behaviors that may lead to poor cybersecurity practices and human vulnerability within enterprises, particularly in healthcare organizations. The RScB

scale has a score range of 0 to 120, with higher values indicating riskier behavior, which is frequently associated with a lack of cybersecurity awareness.

Moreover, in this regard, several machine-learning techniques, such as sequence clustering, can be used to analyze and study user behavior, such as grouping web users with common interests and behaviors. For example, clustering analysis creates a user cluster from web log files. For instance, Facebook’s machine-learning algorithms track every user’s activity on the network to predict their interests, recommend articles, and post notifications on the news feed based on the user’s previous behavior [14].

Fong Tsai showed [16] how collaborative filtering recommendations, which are widely used in recommendation systems on shopping websites, form cluster ensembles. This assumes that people who share the same preferences on certain items also tend to share the same choices on other items. Therefore, clustering based on user logs is done to identify users with similar choices, and it provides recommendations based on the preferences of these “similar neighbors.”

Jiang et al. [17] demonstrated that different machine-learning techniques can be used to extract meaningful data from a huge dataset, including extracting information to analyze user behavior. Callara and Wira [15] suggested an algorithm for user classification based on their dataset and found that it could distinguish 108 groups of users with similar online behavior, which meant they could classify each group with similar behavior as a separate group. They proved that classification techniques are useful in analyzing and labeling test data into known types of classes. Hence, employers can benefit from this classification by providing awareness sessions suited to each class to enhance their employee’s level of security knowledge and keep their assets safe.

Efficient classification techniques have been used by Niranjana and Nitish [11] to enable users to distinguish between phishing and normal websites, classify users as normal users or criminals based on their social media activities (crime profiling), and prevent users from running malicious code by labeling them as “malicious.” However, classifying users into two categories only offers limited options. Concerning the provision of security awareness sessions, a larger number of categories is needed to be more accurate and provide what is needed based on user experience and behavior.

III. METHODOLOGY

The user classification model is a multi-classification problem that aims to classify users into three classes based on the analysis results of their online behavior. To achieve the desired goal of classifying users based on their online behavior and delivering dedicated awareness material to them, a machine-learning-based classification model has been proposed. Assume D , a dataset of website instances, where domain d_i is defined using a set of n features, $F = \{f_1, f_2, \dots, f_n\}$, and each domain $d_i \in D$ is either malicious, suspicious, or normal behavior. The supervised machine-learning algorithm must be trained using D so that the resulting model M can

classify a new domain d_{new} that has not been seen before by M .

The research process has three main phases. The data are collected from users’ records and then prepared using data cleaning and preprocessing. Subsequently, the researchers take various steps to evaluate the classification methods to construct the most effective model of user behavior classification. A diagram describing the workflow of the research procedure is shown in Fig. 1.

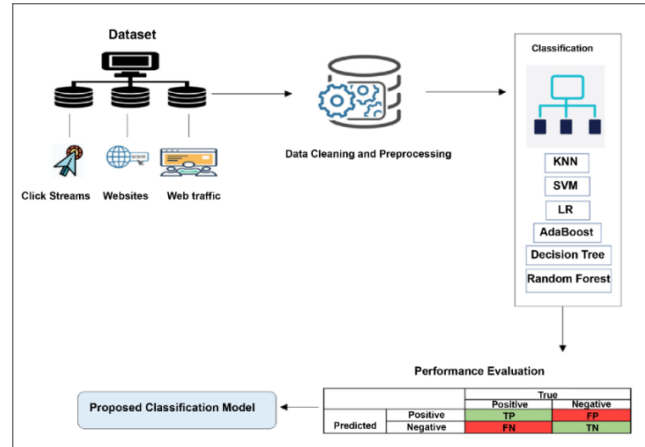


Fig. 1. Security awareness provision based on the user behavior model framework.

A. Data Description

The dataset that was used in this study is from Irvine’s Machine Learning Repository of the University of California [19]. The same dataset was used to investigate and validate the observations. It is an imbalanced multivariate dataset by nature, which has 8,118 instances, each with nine integral forms of attribute characteristics. The data are classified into three user classes to help provide suitable security awareness sessions. These data contain website references/sources which are legitimate or malicious besides the normal references.

Therefore, the dataset contains 8,118 website instances, of which 4,602 are authentic, 2,670 are malicious, and 846 are suspicious. The nine distinct features provided in the dataset that can be used to classify any website as malicious or authentic are server form handlers (SFH), popup window, SSL final state, request URL, URL anchor, web traffic, URL length, domain age, IP address, and the labeled class. They are briefly described in Table I.

B. Model Description

Six well-known classifiers were compared in a supervised learning environment with prior knowledge of the output target set. The classifiers were K nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), adaptive boosting (AdaBoost) classifier, decision tree classifier, and the random forest classifier. Chosen algorithms have been selected as they are commonly used by researchers and in practice for user classification in different fields, as in the work of Kotsiantis *et al.*, [42], Osisanwo *et al.*, [43], and other studies mentioned in this work [11,15,16,17,34,34]. They are described in the following subsections.

TABLE I. DATASET ATTRIBUTES DESCRIPTION

Feature	Description
SFH	SFHs that contain an empty string or “about: blank” are considered doubtful because action should be taken based on the submitted information. In addition, if the domain name in an SFH is different from the domain name of the webpage, then this reveals that the webpage is suspicious because the submitted information is rarely handled by external domains.
Popup window	This is considered a feature, particularly when the website is asking the users to submit any information through a popup window. It is unusual to find a webpage that requests personal information from users via a popup window.
SSLfinal stated	SSL is used to secure communication between a web browser and a web server. This turns a website’s address from HTTP to HTTPS. The “S” stands for “secure.”
Request URL	A request URL examines whether the external objects contained within a webpage, such as images, videos, and sounds, are loaded from another domain.
URL anchor	An anchor is an element defined by the <a> tag. This feature is treated exactly like a “Request URL.”
Website traffic	This feature measures the popularity of the website by determining the number of visitors and the number of pages they visit.
URL length	A URL length can show whether a URL is a suspicious or phishing URL, where specific calculations should be made to determine whether it is a safe URL or a suspicious or phishing URL.
Domain age	Most phishing websites exist for only a short period.
IP address	If an IP address is used that is different from the domain name in the URL, such as “ http://125.98.3.123/fake.html ,” then someone is trying to steal their personal information. Sometimes, the IP address is even transformed into hexadecimal code, as shown in the following link: “ http://0x58.0xCC.0xCA.0x62/2/paypal.ca/index.html .”
Class	This is the class of the domain (malicious behavior = -1, suspicious behavior = 0, and normal behavior = 1).

1) *The KNN classifier*: This is among the most basic classifiers. It works based on a supervised training method, and its technique is based on similarity. The KNN algorithm can perform regression and classification, and it is nonparametric by nature, as it does not make assumptions regarding non-available data. The basic principle is measuring the Euclidean distance from the new point to the nearest previous points, which are the KNNs. The class that has the nearest neighbors is assigned to the given query point.

2) *Support Vector Machine (SVM)*: This machine-learning classification method uses supervised learning, and it is based on the margin or decision boundary, as the SVM selects the optimal margin for classification. However, in this research, we applied the SVM one-vs-Rest (OvR) method of multiclass classification, which was used to create a multiclass SVM classifier. Here, for each class, we created three OvR classifiers. Each classifier should predict a class probability, and the data will be assigned to the highest-probability class.

3) *Logistic Regression (LR)*: This approach works based on the probabilistic prediction of any specified variable and performs the estimation of parameters related to the logistic model. We classified data into more than two classes.

Therefore, we had $y = \{0, 1 \dots n\}$. A one-vs-all strategy was used, in which we trained three distinct binary classifiers, each designed to recognize a specific class. Subsequently, we used these classifiers to predict the correct class.

4) *AdaBoost*: This approach can perform both classification and regression. The working mechanism is based on the meta-estimation and ensemble method. Through this method, weak learning is converted into stronger learning. In the beginning, it uses a basic learning method model and performs repetitive adjustments of the data distribution to increase the accuracy of the next model based on the existing model performance.

5) *Decision tree*: This is a tree-type classifier, and it has nodes, branches, and leaf nodes. The internal nodes are the dataset and features, whereas the branches represent the decision-making rules, and the leaf is the outcome. Thus, it is fundamentally a graphic illustration depicting all possible outcomes of a problem and its conditions. The classification and regression tree algorithm is used to form the tree structure. It is nonparametric by nature and classifies nonlinear data efficiently. It classifies each branch using the decision rules.

6) *Random forest*: The random forest approach is extremely efficient, and its training requires little time. Its accuracy and other performance measures are very high, even when datasets are large or contain missing data. It has parallel decision trees. Thus, it is a type of bagging ensemble. For the classification task, the output is considered the data found at the bottom of the node, while for the implementation of the regression, the mean of all the trees is considered the final output. Let the trees be denoted by $h_1(x), h_2(x), \dots$. AdaBoost, $h_k(x)$; the training data are given as X, Y , and the margin function can be defined as the equation given below:

$$\text{mg}(X, Y) = \text{av}_k I(h_k(X) = y) - \max_{j \neq y} \text{av}_k I(h_k(X) = j) \quad (1)$$

The classification models are implemented on an unbalanced dataset. Each classifier is trained and tested on the dataset.

C. Feature Importance

Feature importance refers to techniques that calculate a score for all the input features for this model—the score represents the importance of each feature. In other words, it indicates strategies for valuing input features depending on their predictive power for a target variable (rank features based on their effect on the model’s prediction.). Feature importance is essential in the context of understanding the data that go into a model, model improvement, or model simplification, which means, in the case of reducing the model dimensionality, high-scoring features could be kept, and the features with the lowest scores could be deleted because they were not necessary.

Because of the points, feature importance scores are a critical component in predictive modeling, as they provide enlightenment of the data and the model. Let D be a dataset of m classes; a represents a feature that takes V possible values $\{a_1, a_2, \dots, a_v\}$ in D . Let D_v be the subset of samples from D that takes the value of a_v for feature a , and let p_i be the

probability that a sample belongs to class i . In the proposed model, the following measures are used: information gain, gain ratio, Gini index, and Pearson product—moment correlation coefficient. These measures have been chosen because they are easy to understand and execute, have light computational requirements, and are frequently successful with various datasets. They are described as follows [19, 20]:

1) *Information gain*: This metric identifies the features that provide the most information about a class and must highlight that entropy plays a crucial role in measuring information gain. Entropy measures the uncertainty of the data. From a different perspective, entropy measures how difficult it is to guess the label of a random sample from a dataset, where low entropy indicates that the data labels are quite uniform, and high entropy indicates that the labels are in confusion [21]. Information gain computes the difference between the entropy before and after a split and specifies class element impurity. The information gain metric investigates the information content of messages; the information gain can be determined by separating dataset D by features, as follows:

$$\text{Gain}(D, a) = \text{Ent}(D) - \sum_{v=1}^V \frac{|Dv|}{|D|} \text{Ent}(Dv) \quad (2)$$

where $\text{Ent}(D)$ is the entropy. By dividing D based on feature a , a high information gain value indicates that the archived data is of greater purity.

2) *Gain ratio*: The gain ratio attempts to reduce the bias of information gain by introducing a normalizing term known as intrinsic information (II). II is the level of difficulty in guessing the branch in which a randomly selected sample is placed. The feature gain ratio is calculated as Gain ratio = information gain/II, which means mathematically:

$$\text{Gain_Ratio}(D, a) = \frac{\text{Gain}(D, a)}{IV(a)}, \quad (3)$$

where $IV(a)$ denotes the intrinsic value of a feature a and is calculated as follows:

$$IV(a) = - \sum_{v=1}^V \frac{|Dv|}{|D|} \log_2 \frac{|Dv|}{|D|} \quad (4)$$

3) *Gini Index*: This is also known as Gini impurity and measures the degree or probability of a variable being incorrectly classified when randomly selected. It measures the dataset impurity. If all the elements in a class belong to a single class, then it can be called pure. In the calculation of impurity, the weight of the feature based on the class label has been calculated. The degree of the Gini index varies between 0 and 1, and a lower Gini index means a higher dataset purity [22]. It can be calculated as follows

$$\text{Gini}_{index}(D, a) = \sum_{v=1}^V \frac{|Dv|}{|D|} \text{Gini}(Dv) \quad (5)$$

where

$$\text{Gini}(D) = 1 - \sum_{i=1}^m P_i^2 \quad (6)$$

4) *Mattheus correlation coefficient*: Brian W. Mattheus developed the Mattheus correlation coefficient (MCC) in 1975 using Karl Pearson's phi coefficient, and it has become a widely used metric for evaluating the effectiveness of machine-learning techniques, with extensions for multiclass cases [23]. It has a value range between [-1 and 1] that measures the strength and direction of the relationship between two variables as a strong correlation, no correlation, or an inverse relationship.

5) *Kappa*: Cohen's kappa builds on the idea of measuring the concordance between the predicted and true labels, which are regarded as two random categorical variables [24]. Two categorical variables can be compared by constructing a confusion matrix and determining the marginal row and column distributions. Therefore, we can begin using Cohen's kappa indicators as ratings of the dependence (or independence) between the model's prediction and actual classification.

In the multiclass case, the calculation of Cohen's kappa score is as follows [25]:

$$K = \frac{c \times s - \sum_k^K p_k \times t_k}{s^2 - \sum_k^K p_k \times t_k}, \quad (7)$$

where

- $C = \sum_k^K C_{kk}$ the total number of elements correctly predicted
- $S = \sum_i^K \sum_j^K C_{ij}$ the total number of elements
- $p_k = \sum_j^K C_{kj}$ the number of times that class k was predicted (column total)
- $t_k = \sum_j^K C_{ki}$ the number of times that class k was predicted (rows total)

D. Data Preprocessing

The original data must be preprocessed to remove irrelevant and redundant log entries. The following preprocessing techniques were applied to the collected data before they were trained and analyzed. Each technique is described next.

1) *Check and remove null or missing entries*: This step is considered one of the most essential steps in data cleaning. All missing data are identified and then removed. It should also clean the data of all irrelevant information, such as "Nan," "n/a," or any other irrelevant values having a number in the URL attribute. These are removed using Python Regex. Empty entries are removed as well.

2) *Data normalization and standardization*: The process in which the data is cleaned is known as data normalization. This cleaning makes the data regular for all the values of features, which leads to improved segmentation. It removes all the unstructured and redundant data to provide logical data storage. This type of data management is considered particularly crucial for large databases. The raw data hinder the achievement of high efficiency. This problem is dealt with

through data normalization. All the feature values are compressed between [0, 1]. The mean is shifted to 0, and the standard deviation is maintained at 1, so the data can be standardized and easily manipulated. Most machine-learning algorithms display noticeable increments in efficiency after the implementation of normalization.

E. Principal Component Analysis (PCA)

Essentially, the algorithm follows the data relationships to a base field and then sequentially applies mathematical functions in the data in different columns and rows along this path to generate the final feature [26]. The performance of classification algorithms may be compromised because of redundant or highly correlated features. Thus, we implemented dimensionality reduction using PCA, as it reduces the size of the feature space while retaining a significant amount of the information [27]. In this regard, many studies have indicated that PCA is less noise-sensitive than other dimension reduction methods [28, 7].

F. Experimental Setting

The proposed model was implemented using Python and Pandas library. A personal computer was used for this experiment, with the following specifications: operating system: macOS Monterey; chip: Apple M1 Pro; total number of cores (processors): eight (six performance and two efficiency); and OS Loader, version: 7459.141.1. In addition, the programming language Python was used.

Based on the parameter settings, the performance of various algorithms can vary. In this work, the algorithms were run using the following parameters:

1) *KNN model classifier*: $K = 5$, weights = “uniform”, algorithm = “auto” “fit method is `model1.fit(X_train,y_train)`, leaf_size = 30, $p = 2$ (Euclidean distance), metric = “minkowski”.

2) *SVM classifier*: The regularization parameter is set to 1, with a linear kernel, no class weights, and a shrinking heuristic.

3) *LR classifier*: The norm of the penalty = L2. No class weights, fit intercept is set to true, maximum iterations = 100, and for multi_class = “auto”.

4) *AdaBoost classifier*: integer value = 42.

5) *Decision tree classifier*: Decision tree classifier (random_state = 42) with no maximum depth, which means nodes are expanded until all leaves are pure or until all leaves contain less than min_samples_split samples, and the splitter is the “best”.

6) *Random forest classifier*: One hundred trees, with no maximum depth and a minimum number of splits = 2.

The experiments were designed using different machine-learning and data-analytics libraries, including scikit-learn [29], Numpy [14], and Pandas [31]. Six machine-learning algorithms (described previously) were employed along with the PCA-based feature importance measure with reduced dimensions. Standard 10-fold cross-validation [32] train/test trials were run by partitioning/splitting the entire dataset into training and testing (proportions of 70% and 30%). We

ensured that the test data contained a fair distribution for all classes. The following experiments were designed with consistent classifier configurations:

1) *Train* and test the seven machine-learning algorithms over the individual datasets.

2) *Train* and test the five machine-learning algorithms over the PCA-based dimension-reduced datasets using a 10-fold CV to compare the performances.

G. Performance Measures

After performing classification, its performance and results must be gauged without specific markers. Therefore, to evaluate a classifier’s capabilities, various performance measures can represent the classification quality of different classifiers on any given data. This provides a deeper insight into the classification techniques’ efficiency than that which using basic accuracy percentages can achieve. The performance evaluation is accomplished using performance metrics such as confusion matrix, precision, recall, and F1 score, as well as basic accuracy. Brief descriptions of each of the performance measures are as follows:

1) *The confusion matrix* represents the relationship between the actual and predicted values. The following briefly describes the confusion matrix with its four basic elements:

2) *True Positive (TP)*: A vector that gives a count of correctly classified data (presence of condition). Mathematically, this can be calculated by $TP/(TP+FP)$.

3) *False Positive (FP)*: A vector that gives the incorrect classification of data (e.g., the detection of a condition that is not present). Mathematically, this can be calculated by $TN/(TN+FN)$.

4) *True Negative (TN)*: A vector that shows the number of correctly classified data that do not possess the condition (absence of condition).

5) *False Negative (FN)*: A vector that gives the count of wrongly classified data (detected the absence of a condition when it was present).

a) *Accuracy*: The most basic and extensively relied upon measurement is accuracy, as calculated in Eq. 8 below. It represents the accuracy of the classification results and is the fraction or percentage of a classifier’s total correct identifications against the classifier’s total outcomes, both correct and incorrect.

$$\text{Accuracy} = \frac{\text{Correctly classified samples}}{\text{total number of classifications}} \quad (8)$$

b) *Precision*: This measurement tells us how precise the classifier results are. It gives the percentage of correctly identified positive outcomes against total positive outcomes, which includes false positives.

c) *Recall*: Recall measures the sensitivity of the classifier. It gives the recognition rate of a classifier. A recall is the proportion of correct positive outcomes against the total number of actual positives present in the dataset. Therefore, it includes false negatives.

d) *F1 score*: The F1 measurement is an amalgamation of both precision and recall. It is essentially the subjective average of both, namely the recall and precision values. It provides more precise estimations of incorrect outcomes than accuracy when the dataset is imbalanced.

e) *Receiver operating characteristic (ROC) area*: The ROC metric is used to evaluate the quality of multiclass classifiers. The true positive rate is typically plotted on the Y axis and the false positive rate (FPR) on the X axis. For multiclass problems, ROC curves can be plotted by comparing one class against the others. Applying this OvR to each class will give results in the same number of curves as classes. The ROC score can also be calculated separately for each class. ROC values range between 0 and 1. A model with 100% incorrect predictions has a value of 0.0 while one with 100% accurate predictions has a value of 1.0.

f) *Precision-recall curve (PRC) area*: PRC can be referred to as the relationship between precision and recall (sensitivity) and is regarded as a more suitable metric for unbalanced datasets. PRC can be calculated by integrating the piecewise function. Consequently, the PRC tends to intersect significantly more frequently than the ROC. The primary distinction between the two is that the number of true negative results is not factored into the PRC because the precision-recall curves are only affected by true positives in most cases. The PRC is generally a tortuous curve, fluctuating upwards and downwards [33].

IV. MODEL RESULTS

In this model, we were looking to classify users into three classes using each of the six best classifiers regarding the performance measurements that were applied to their evaluation and selection. Each classifier was trained and tested separately to evaluate it in a different portion of the dataset for each classification model with different testing options. We had 70% of the dataset for training and 30% for testing the model besides applying PCA to the dataset. In addition, we performed cross-validation to improve the effectiveness and accuracy of the classification.

A. ML Classification Results

The following Table II and Table III, illustrate the classification performance of the six classifiers used in this work. The tables show the evaluation measure for all six classification models trained on 70% of the dataset and tested on 30%.

TABLE II. PERFORMANCE MEASURES FOR THE SIX CLASSIFICATION MODELS TRAINED ON 70% OF THE DATASET

Classifier	KNN	SVM	LR	AdaBoost	Decision tree	Random forest
Accuracy	93.54%	88.21%	86%	96.07%	95.40%	96.58%
Recall	93.5%	88%	86%	96.1%	95.5%	96.6%
Precision	93.6%	87%	85%	96.1%	95.6%	96.6%
F1 measures	93.6%	87%	85%	96.3%	95.5%	96.6%
MCC	88%	80%	76%	93%	91%	94%
Time (seconds)	1.25	0.01	0.03	0.02	0	0.13

TABLE III. PERFORMANCE MEASURES FOR THE SIX CLASSIFICATION MODELS TESTED ON 30% OF THE DATASET

Classifier	KNN	SVM	LR	AdaBoost	Decision tree	Random forest
Accuracy	92.89%	88.%	87%	95.8%	94.62%	96.09%
Recall	92.8%	88.8%	87%	95%	94.6%	96.1%
Precision	93%	88.4%	85%	95%	94.7%	96.1%
F1 measures	93%	88.4%	86%	95%	94.6%	96.1%
MCC	87%	81%	77%	92%	91%	93%
Time (seconds)	0.56	0.01	0.01	0.01	0	0.05

As presented in the tables previously, we can see that all the classifiers have been applied to evaluate each classifier's performance. Training data helps construct a machine-learning model and teaches it what the expected outcomes should look like, while the model examines the dataset repeatedly to understand its characteristics and optimize its performance. In contrast, after a machine-learning model is constructed using the training dataset, it must be tested to evaluate the performance of each classifier to select the optimal classifier from those included.

Table II and Table III show the training and testing results regarding the performance matrix evaluation. Comparing the results of all classifiers using part of the dataset, the final results show that the best accuracy is for the random forest classifier, although some of the classifiers, such as AdaBoost and decision tree, have results close to the random forest classifier. Additionally, the LR classifier achieves the lowest accuracy value in both testing and training the model compared to the other classifier models. In this study, AdaBoost was a combination of J48 and decision tree, where the J48 algorithm is closer to the random tree algorithm even in the time it requires for execution. J48 is an algorithm that C4 (one of the decision tree classifiers) employs to generate a decision tree (an extension of ID3). Also referred to as a statistical classifier [30], the J48 algorithm is used to classify various applications and produce accurate classification results, to produce more accurate and fairer comparison results.

The random forest algorithm has the highest accuracy but requires significantly more time to generate a model than the decision tree and AdaBoost algorithms. Besides measuring each classifier's accuracy, because we have an imbalanced dataset, another measurement could assist us in deciding which classifier would perform the best and enable us to have more accurate evaluation results.

We also considered MCC because this indicator is viewed as an effective solution to overcoming the class imbalance issue [34]. In the evaluation, we also considered the F1 measurement, as it is widely used in most application areas of ML, particularly in multiclass cases [35]. Because we had close results for accuracy and time for some of the classifiers, for additional evaluation indicators, we added MCC results to the previously presented tables as well as included them and the F1 results in selecting the best classifier for this proposed

model. The random forest classifier has the highest MCC and F1 result among all the machine-learning classifiers.

B. Feature Importance

Next, to understand to what degree each feature contributes to model prediction, which will affect its performance and accuracy in the model, we analyzed feature importance using the four-feature importance measures. The following Fig. 2 shows each feature’s rank and score; the scores represent the “importance” of each feature. A higher score indicates that the feature will have more impact on the model used to predict a particular variable.

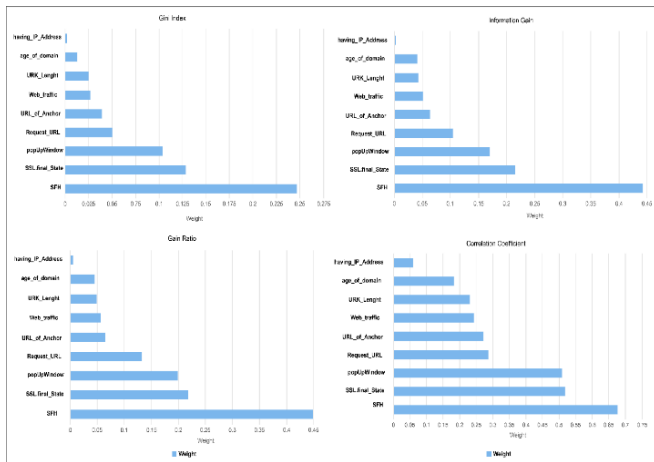


Fig. 2. Top four features and their corresponding weights using information gain, gain ratio, Gini index, and correlation coefficient.

The results show that the top four features are SFH, SSL_final_state, popup window, and requested_URL.

C. PCA

PCA, in this context, is the concept of reducing the number of variables of the dataset while retaining as much information as possible. Accuracy naturally suffers when a dataset’s variables are reduced, but the aim of dimensionality reduction is to sacrifice a little accuracy in return for greater simplicity because machine-learning algorithms can analyze data much quickly and easily with smaller data sets as there are fewer extraneous variables to process. Table IV and Table V. show the results of applying PCA to each classification model.

TABLE IV. PERFORMANCE MEASURES WITH PCA (70% TRAINING DATASET)

Classifier	KNN	SVM	LR	AdaBoost	Decision tree	Random forest
Accuracy	95%	86%	85%	97%	95%	98%
Recall	95%	86%	85%	97%	95%	98%
Precision	95%	86%	85%	97%	96%	98%
F1 measures	95%	86%	85%	97%	95%	98%
MCC	93%	79%	77%	95%	93%	96%
Time (seconds)	1.27	0.42	0	0.56	0	0.19

TABLE V. PERFORMANCE MEASURES WITH PCA (30% TESTING DATASET)

Classifier	KNN	SVM	LR	AdaBoost	Decision tree	Random forest
Accuracy	93%	85%	85.18%	95%	93%	96%
Recall	93%	85%	85%	95%	94%	96%
Precision	93%	90%	89%	95%	94%	96%
F1 measures	93%	86%	86%	95%	94%	96%
MCC	87%	79%	78%	92%	89%	93%
Time (seconds)	0.58	0.2	0.1	0.10	0.03	0.07

According to the previous presented tables (Table IV and Table V), we can see the improvement in accuracy when the PCA was applied to the dataset because of dimensionality reduction where the redundant and irrelevant data have been removed; in other words, the data that have no significant effect on the classification results have been removed. Additionally, the improvement in the MCC results is noticeable.

For further investigation, and as the final results of all six classifiers were similar, a 10-fold data split was constructed, as shown in the following Fig. 3, to understand how the algorithms performed.

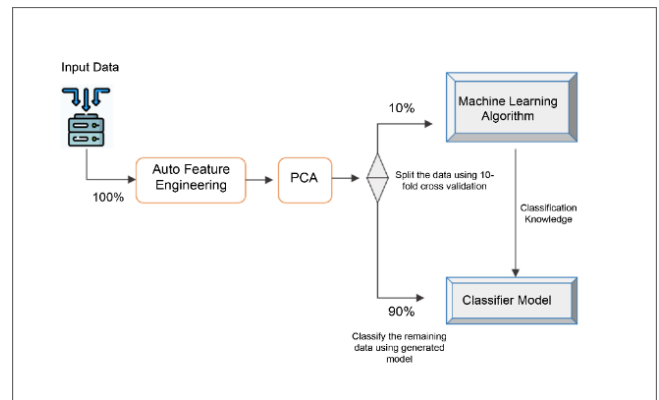


Fig. 3. Cross-validation process model.

The classifier constructed nine identical instances of the dataset and then split the data in each of these instances into 10% for training and 90% for testing. Each of these nine instances was trained/tested with a unique split. Finally, the result from each of these instances was combined into a final result. Because nine combinations of 10% of the data were used to classify the data, a reasonably realistic result could be obtained using this 10-fold cross-validation split.

Using cross-validation emphasizes that, as previous Fig. 4 and Fig. 5 shown, although all the classifier results are similar to each other, the random forest classifier shows the best performance regarding all performance measures and, in particular, the lowest FPR (2.2%), with incorrectly classified instances of 4% in the cross-validation test.

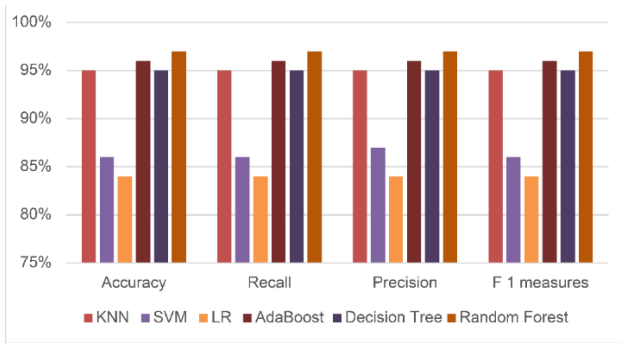


Fig. 4. Cross-validation.

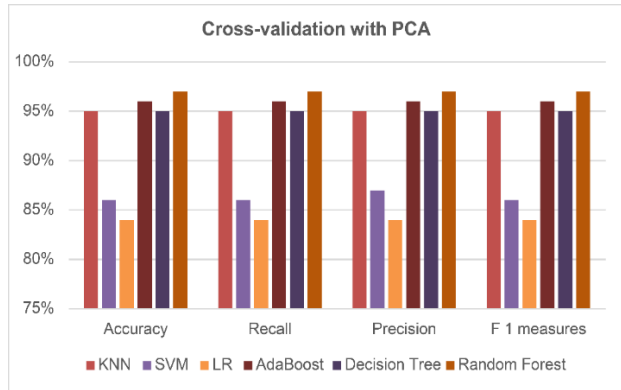


Fig. 5. Cross-validation with PCA.

D. Proposed Model

This study aimed to propose a model that can assist organizations in providing dedicated and targeted cybersecurity awareness sessions to their employees based on an analysis of their online behavior. The problem at hand was formulated as a multiclass problem. We differentiated between three classes: malicious, suspicious, and normal. Based on influential features and the best-performing classifier we identified, we propose an ML-based classification model. Fig. 6 shows the proposed model.

The dataset was first fed into the classifier, which was then used to extract features. Following that, a few preprocessing techniques were applied to ensure that the dataset was clean. After that, we applied the machine-learning classification models to the dataset.

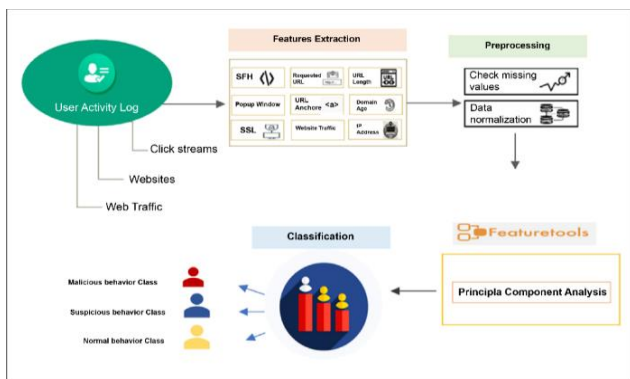


Fig. 6. Machine-learning-based classification model.

The Kruskal–Wallis test was used to compare the performance of the various models in the study. The Kruskal–Wallis test is a nonparametric statistical test that is computed based on the rank and the sum of ranks. The null hypothesis assumes that the performance measures of the models are drawn from the same distribution and that any differences are due to chance.

The hypothesis of the test is given below:

H_0 : The performances of the models are equal (i.e., there are no statistically significant differences in model performances).

H_A : At least one model performance is different (i.e., there are statistically significant differences in model performances).

1) *Test statistics*: The test statistic of the Kruskal–Wallis, H measures the differences among the performance of the groups and is given by the following:

$$H = \frac{12}{N(N+1)} \left(\sum \frac{R_i^2}{n_i} \right) - 3(N+1)$$

where n_i total number of observations in the model i
 R_i the sum of the ranks of model i
 N the total number of observations across all models.

The Kruskal–Wallis test statistic approximates a chi-square distribution with $k-1$ degrees of freedom, where k is the number of groups (models).

The observations for the test are obtained from the classification accuracy of each of the models from 10-fold cross-validation. Hence, this ensures that each classifier used is evaluated on the same splits of the dataset via the 10-fold cross-validation. These observations (classification accuracies from the 10-fold cross-validation) are provided in the appendix below. The Kruskal–Wallis test is then used to compare whether there is a statistically significant difference among the performance of these models. All analyses were implemented using Python software.

2) *Test results*: The test statistics and the associated p -values are given below in Table VI.

TABLE VI. TEST STATISTICS AND THE ASSOCIATED P-VALUES

	H statistics	p -value
Training set	45.4512	1.1745×10^{-8}
Testing set	48.9979	2.2215×10^{-9}
PCA with Training set	49.0362	2.1818×10^{-9}
PCA with Testing set	49.3816	1.8544×10^{-9}

Decision Rule

Reject H_0 if the p -value ≤ 0.05 ; else, fail to reject H_0 .

Because the p -value associated with any of the H statistics is less than 0.05, we reject H_0 . Hence, enough evidence

supports the alternative hypothesis that at least one of the model performances is different. Therefore, there are statistically significant differences in model performances.

3) *Post hoc statistical test: Dunn’s Test with the Holm–Bonferroni Correction:* Given that the Kruskal–Wallis test showed evidence of statistically significant differences in model performance, the Dunn’s test with Holm–Bonferroni p-value correction was conducted to ascertain which pairs of models differ significantly from each other.

Dunn’s test is a nonparametric pairwise post hoc test used to compute the rank-based Z-values for pairs of the models and convert these values into p-values. The Holm–Bonferroni correction is then applied to these p-values to control for family-wise error rate (FWER). FWER refers to the probability of committing at least one type I error among the pairs of comparisons. All computations are conducted using Python.

4) *Hypothesis:* The hypothesis of this test for each of the pairs of models is as follows:

H_0 : There is no statistically significant difference between the pair of models compared.

H_A : There is a statistically significant difference between the pair of models compared.

5) *Decision rule:* Reject H_0 if the Holm–Bonferroni Adjusted p-value ≤ 0.05 ; else fail to reject H_0

TABLE VII. RESULT OF DUNN’S TEST WITH THE HOLM–BONFERRONI CORRECTION ON TRAIN PERFORMANCE

Model 1	Model 2	HB Adj. p-value	Result
KNN	SVM	1	Not significant
KNN	LR	0.263806	Not significant
KNN	AdaBoost	1	Not significant
KNN	Decision Tree	0.899874	Not significant
KNN	Random Forest	1	Not significant
SVM	LR	0.899874	Not significant
SVM	AdaBoost	1	Not significant
SVM	Decision Tree	0.263806	Not significant
SVM	Random Forest	1	Not significant
LR	AdaBoost	1	Not significant
LR	Decision Tree	0.000080	Significant
LR	Random Forest	0.935495	Not significant
AdaBoost	Decision Tree	0.002421	Significant
AdaBoost	Random Forest	1	Not significant
Decision Tree	Random Forest	0.218907	Not significant

TABLE VIII. RESULT OF DUNN’S TEST WITH THE HOLM–BONFERRONI CORRECTION ON TEST PERFORMANCE

Model 1	Model 2	HB Adj. p-value	Result
KNN	SVM	1	Not significant
KNN	LR	1	Not significant
KNN	AdaBoost	1	Not significant
KNN	Decision Tree	1	Not significant
KNN	Random Forest	0.003238	Significant
SVM	LR	0.004285	Significant
SVM	AdaBoost	1	Not significant
SVM	Decision Tree	0.004285	Significant
SVM	Random Forest	1	Not significant
LR	AdaBoost	0.211045	Not significant
LR	Decision Tree	1	Not significant
LR	Random Forest	0.0000003	Significant
AdaBoost	Decision Tree	0.211045	Not significant
AdaBoost	Random Forest	0.045905	Significant
Decision Tree	Random Forest	0.0000003	Significant

TABLE IX. RESULT OF DUNN’S TEST WITH THE HOLM–BONFERRONI CORRECTION ON TRAIN PCA

Model 1	Model 2	HB Adj. p-value	Result
KNN	SVM	1	Not significant
KNN	LR	1	Not significant
KNN	AdaBoost	1	Not significant
KNN	Decision Tree	0.031372	Significant
KNN	Random Forest	1	Not significant
SVM	LR	1	Not significant
SVM	AdaBoost	1	Not significant
SVM	Decision Tree	0.092881	Not significant
SVM	Random Forest	1	Not significant
LR	AdaBoost	1	Not significant
LR	Decision Tree	0.018019	Significant
LR	Random Forest	1	Not significant
AdaBoost	Decision Tree	0.001791	Significant
AdaBoost	Random Forest	1	Not significant
Decision Tree	Random Forest	0.62984	Not significant

TABLE X. RESULT OF DUNN’S TEST WITH THE HOLM–BONFERRONI CORRECTION ON TEST PCA

Model 1	Model 2	HB Adj. p-value	Result
KNN	SVM	0.056299	Not significant
KNN	LR	0.692211	Not significant
KNN	AdaBoost	1	Not significant
KNN	Decision Tree	1	Not significant
KNN	Random Forest	0.007251	Significant
SVM	LR	0.0000086	Significant
SVM	AdaBoost	1	Not significant
SVM	Decision Tree	0.000714	Significant
SVM	Random Forest	1	Not significant
LR	AdaBoost	0.001994	Significant
LR	Decision Tree	1	Not significant
LR	Random Forest	0.0000003	Significant
AdaBoost	Decision Tree	0.056299	Not significant
AdaBoost	Random Forest	1	Not significant
Decision Tree	Random Forest	0.000047	Significant

The results above show that there exists at least one instance where pair of models are statistically different regarding performance.

V. DISCUSSION

Currently, user behavior is one of the most critical factors in organizations' cybersecurity, and it can put the organization's safety, data, assets, reputation, and individuals at risk. Thus, providing cybersecurity training for users or employees plays a vital role in improving their attitude and behavior when online, particularly when the training is directed and targeted based on user needs and deficiencies.

Due to the large number of attributes and high volume of online data, we employed machine-learning techniques in the context of providing cybersecurity awareness by analyzing online user behavior. In this context, the main objective was the enhancement of people's cybersecurity awareness through the provision of targeted cybersecurity awareness programs that would lead to a decrease in cybersecurity issues and intrusions inside an organization.

Although user behavior analysis and the use of machine-learning techniques for analyzing user behavior are not new, the novelty of this paper lies in the fact that it is among the first few research that analyses human online behavior and applies ML to target employees with suitable awareness materials, the primary objective of this study differed from those of previous models and other studies. The concept of user behavioral analysis has been included previously in a number of fields and domains, such as marketing applications, to adopt new and efficient marketing strategies that are based on user data (i.e., utilizing recorded information of the past activities of potential clients in data-based behavioral marketing) [36]. It has also been included in recommendation systems by predicting user interests from a user's last browsing and searching activities, for example, by recommending specific articles for readers or an item of clothing during shopping [34].

Moreover, ML is used to classify users, such as on social media. It can be applied to building a practical system for detecting fake identities by using server-side clickstream models to group users with similar clickstreams into clusters or analyze user browsing behavior on specific websites [35], including e-commerce, education, and healthcare. The aim is the personalization or targeting of users with advertisements based on their browsing behavior. Thus, the application of machine-learning techniques helps classify users with a high degree of accuracy. In the security domain, its value has been proven in the fight against fraud and other applications [37]. Moreover, ML is used in the detection of phishing emails using algorithms. This can automate the detection of phishing emails using a variety of techniques, including deep-learning detectors that automate the process [38], where deep-learning algorithms have produced impressive results with unstructured data such as email data [39].

This proposed model can aid organizations in maintaining the security of their assets and data, as we include the human factor by enhancing the awareness levels of their employees regarding cybersecurity threats by providing appropriate

training and awareness based on the analysis of their online behavior that may help the organization in classifying users based on the analysis results.

Ryu et al. [18] and many others demonstrated the importance of personal security factors in this area. They showed the significance of raising awareness of the importance of security in industries. As a result, regardless of the type of security system in place, considering the importance of employees' online awareness and behaviors is critical.

Many other researchers [34–36] have shown that a strong awareness-raising program is required to ensure that employees understand their respective IT security duties and roles to protect the IT resources delegated to them. However, these studies achieved low accuracy in measuring users' online awareness; for example, questionnaires or surveys were published to a general audience, and the analysis was performed based on their answers [33]. This approach fails to analyze employees' actual online behavior that reflects their cybersecurity knowledge. As a result, the awareness content that is subsequently provided is not suitable for each individual.

In this study, we applied several machine-learning classifications to the same dataset with the same percentage split: 70% for training the model and 30% for testing the model. Thereafter, we compared the final results of the performance measures among all classifiers to determine the best one. The results demonstrated that the random forest classifier was the best option to choose with the best results, and it could be applied for analyzing user behavior inside the organization. Random forest achieved the highest accuracy rate in both training and testing sets of the whole dataset with different methods of testing and different measures that have been used, which are the accuracy, MCC, and F1 measures.

For the AdaBoost, decision tree, and random forest classifiers, the accuracy rates were similar. Therefore, we included the MCC and F1 measurements to ensure a more accurate comparison, rather than just taking into consideration the FPR and which classifier had the lowest FPR. PCA was also applied to the concept of reducing the dimensionality of the dataset used in the model, and cross-validation was used to validate each classifier.

Theoretically, when considering the computational costs of the random forest classifier, the complexity of the test time of a random forest of size T , which is the number of trees to build, and the maximum depth D is $O(T \cdot D)$, which is 0 by default and is the unlimited depth of the tree. Another important disadvantage is the memory space required for random forest classification, which is calculated by $O(2^D)$ [33]. This experiment showed that the running time to build the model is 0.23 s, on average, and the time required to test the model on 5,683 instances of training data is 0.11 s. Additionally, the time required to build the model is 0.19 s, and the time required to test the model on the supplied test set is 0.09 s for 2,435 instances.

Random forest showed its effectiveness in the classification process, as it did in many previous works, such

as in Android malware classification [40], where it performed very well with an accuracy of over 99%. In general, the samples were correctly classified, and the highest number of misclassified cases resulted from samples from the malicious class being mistakenly assigned to the benign class.

Moreover, Farnaaz and Jaber [41] used random forest classification to detect intrusions on a system, where the random forest classifier was used to classify four types of attacks. According to empirical findings, the proposed model was effective, with a low false alarm rate and a high detection rate.

Thus, the experimental results conclude that users can be successfully classified based on their online behavior to target them with the correct awareness materials using a machine-learning-based model.

VI. CONCLUSION

The causes of and methods for preventing security issues and risks to any organization are continually changing as a direct result of the ongoing evolution of cybersecurity threats. In addition, individuals' knowledge levels, technical skills, and levels of awareness regarding cybersecurity vary, which is one of the reasons for the difficulty in controlling their online behavior and the associated risks. Because of this, the measurement and analysis of online behavior are now absolutely necessary for any organization that wants to protect its assets from both internal and external breaches of security. A substantial number of earlier studies have established a clear connection between online users' actions and various problems and dangers related to cybersecurity. Regardless of the security technology in place, the most reliable indicator of potential vulnerabilities in an organization or network is users' actions when they are online. Providing directed and dedicated awareness sessions and training regarding cybersecurity is essential in any organization, and this must be managed appropriately.

In this study, we proposed a machine-learning-based model that can assist organizations in providing targeted awareness sessions to their employees based on an analysis of the employees' behaviors. The model will classify the users into three classes: malicious, suspicious, and normal behavior. This classification will ultimately increase awareness of particular behaviors. It may enable organizations to target each employee segment with appropriate sessions and training, increasing the effectiveness of resources.

To achieve this objective, a machine-learning model can be applied to identify patterns in users' web activities and, as a result, classify users according to their activities in virtual spaces. The primary goal of the proposed model is to help organizations target users with sessions of security awareness that are specific and tailored to their needs. Raising awareness can be automated based on specific behaviors, which may result in an effective process that saves organizations time and money. Six well-known machine-learning algorithms, namely KNN, LR, SVM, AdaBoost, decision tree, and random forest classifiers, were trained and tested independently on a user behavior records dataset by splitting the dataset into a 70% training dataset and a 30% testing dataset. The random forest

classifier showed superior performance among all the classifiers regarding the accuracy, F-measure, and the MCC measure. While applying PCA, the model also demonstrates a high accuracy rate, low FPR, high recall, and precision, as well as high F-measures.

Furthermore, as this model is based on machine learning, Machine learning methods at some point also have limitations, as when applied to security that can result in amplified nuances. They can give false positives and false negatives, causing them to miss detection, or insiders can corrupt the dataset, which will lead to wrong outcomes or corruption of the model itself. Furthermore, hackers are also learning machine learning and applying them to their hacking procedures and fishing for loopholes to exploit.

This model has the potential to undergo further development by automatically learning user classes to set up appropriate awareness sessions and training without human intervention. In subsequent research, an improved feature analysis might be included with the goal of making the model more precise. Another potential development would be the incorporation of additional user behavior categories. In addition, a monitoring strategy can be used to observe user behavior. Management can be notified if there is no change in the manner in which users conduct themselves while online. In the future, we plan to increase the number of classes for classifying users and the amount of automated content to be sent to each class to enhance the model's value to organizations.

REFERENCES

- [1] Chen, C. C., Shaw, R. S., & Yang, S. C. (2006). Mitigating information security risks by increasing user security awareness: A case study of an information security awareness system. *Information Technology, Learning & Performance Journal*, 24(1).
- [2] Johnston, A. C., Warkentin, M., McBride, M., & Carter, L. (2016). Dispositional and situational factors: influences on information security policy violations. *European Journal of Information Systems*, 25(3), 231-251.
- [3] Donalds, C., & Osei-Bryson, K. M. (2020). Cybersecurity compliance behavior: Exploring the influences of individual decision style and other antecedents. *International Journal of Information Management*, 51, 102056.
- [4] Bishop, M. (2005). Authentication. In *Introduction to Computer Security* (pp. 171-96). Addison-Wesley.
- [5] de Zafra, D. E., Pitcher, S. I., Tressler, J. D., & Ippolito, J. B. (1998). Information technology security training requirements: A role-and performance-based model. *NIST Special publication*, 800(16), 800-16.
- [6] Kruger, H. A., & Kearney, W. D. (2006). A prototype for assessing information security awareness. *Computers & security*, 25(4), 289-296.
- [7] Carblanc, A., & Moers, S. (2003). Towards a culture of online security: making information systems trustworthy is a job that concerns everyone. What can be done?. *OECD Observer*, (240-241), 30-32.
- [8] Ryu, S., Kang, Y. J., & Lee, H. (2018, February). A study on detection of anomaly behavior in automation industry. In *2018 20th International Conference on Advanced Communication Technology (ICACT)* (pp. 377-380). IEEE.
- [9] Curme, C., Preis, T., Stanley, H. E., & Moat, H. S. (2014). Quantifying the semantics of search behavior before stock market moves. *Proceedings of the National Academy of Sciences*, 111(32), 11600-11605.
- [10] Bernaschina, C., Brambilla, M., Mauri, A., & Umuhoza, E. (2017). A big data analysis framework for model-based web user behavior analytics. In *Web Engineering: 17th International Conference, ICWE*

- 2017, Rome, Italy, June 5-8, 2017, Proceedings 17 (pp. 98-114). Springer International Publishing.
- [11] Niranjana, A., Nitish, A., Deepa Shenoy, P., & Venugopal, K. R. (2016). Security in data mining-a comprehensive survey. *Global Journal of Computer Science and Technology*, 16(5).
- [12] Kumar, S., & Singh, M. (2018). Big data analytics for healthcare industry: impact, applications, and tools. *Big data mining and analytics*, 2(1), 48-57.
- [13] Robila, S. A., & Ragucci, J. W. (2006). Don't be a phish: steps in user education. *Acm sigse bulletin*, 38(3), 237-241.
- [14] Nunes, P., Antunes, M., & Silva, C. (2021). Evaluating cybersecurity attitudes and behaviors in Portuguese healthcare institutions. *Procedia Computer Science*, 181, 173-181.
- [15] Callara, M., & Wira, P. (2018, November). User behavior analysis with machine learning techniques in cloud computing architectures. In 2018 International Conference on Applied Smart Systems (ICASS) (pp. 1-6). IEEE.
- [16] Tsai, F. S. (2010). Comparative study of dimensionality reduction techniques for data visualization. *Journal of artificial intelligence*, 3(3), 119-134.
- [17] Jiang, H., He, M., Xi, Y., & Zeng, J. (2021). Machine-learning-based user position prediction and behavior analysis for location services. *Information*, 12(5), 180.
- [18] Ryu, S., Kang, Y.J., & Lee, H. (2018, February). A study on detection of anomaly behavior in automation industry. In 2018 20th International Conference on Advanced Communication Technology (ICACT) (pp. 377-380). IEEE.
- [19] <https://archive.ics.uci.edu/ml/datasets/website+phishing>
- [20] Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16-28.
- [21] Alhogail, A., Al-Turaiki, I.: Improved detection of malicious domain names using gradient boosted machines and feature engineering. *Inf. Technol. Control* 51, 313-331 (2022)
- [22] <https://www.javatpoint.com/entropy-in-machine-learning>
- [23] Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 1-13.
- [24] Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research*, 8(4), 187.
- [25] Tallón-Ballesteros, A. J., & Riquelme, J. C. (2014). Data mining methods applied to a digital forensics task for supervised machine learning. *Computational intelligence in digital forensics: forensic investigation and applications*, 413-428.
- [26] Al-Turaiki, I., & Altwaijry, N. (2021). A convolutional neural network for improved anomaly-based network intrusion detection. *Big Data*, 9(3), 233-252.
- [27] Kanter, J. M., & Veeramachaneni, K. (2015, October). Deep feature synthesis: Towards automating data science endeavors. In 2015 IEEE international conference on data science and advanced analytics (DSAA) (pp. 1-10). IEEE.
- [28] Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202.
- [29] Robila, S.A., & Ragucci, J. W. (2006). Don't be a phish: steps in user education. *Acm sigse bulletin*, 38(3), 237-241.
- [30] Patel, B. R., & Rana, K. K. (2014). A survey on decision tree algorithm for classification. *International Journal of Engineering Development and Research*, 2(1), 1-5.
- [31] Agrawal, R., & Srikant, R. (1995, March). Mining sequential patterns. In Proceedings of the eleventh international conference on data engineering (pp. 3-14). IEEE.
- [32] Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H., & Zhao, B. Y. (2013). You are how you click: Clickstream analysis for sybil detection. In 22nd USENIX Security Symposium (USENIX Security 13) (pp. 241-256).
- [33] Solé, X., Ramisa, A., & Torras, C. (2014). Evaluation of random forests on large-scale classification problems using a bag-of-visual-words representation. In *Artificial Intelligence Research and Development* (pp. 273-276). IOS Press.
- [34] Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2), 442-451.
- [35] Tsoumakas, G., Katakis, I., & Vlahavas, I. (2010). Random k-labelsets for multilabel classification. *IEEE transactions on knowledge and data engineering*, 23(7), 1079-1089.
- [36] Foxall, G. R. (1994). Behavior analysis and consumer psychology. *Journal of Economic Psychology*, 15(1), 5-91.
- [37] Baig, A. R., & Jabeen, H. (2016). Big data analytics for behavior monitoring of students. *Procedia Computer Science*, 82, 43-48.
- [38] Alhogail, A., & Alsabih, A. (2021). Applying machine learning and natural language processing to detect phishing email. *Computers & Security*, 110, 102414.
- [39] Halgaš, L., Agraftotis, I., & Nurse, J. R. (2020). Catching the phish: Detecting phishing attacks using recurrent neural networks (rnns). In *Information Security Applications: 20th International Conference, WISA 2019, Jeju Island, South Korea, August 21-24, 2019, Revised Selected Papers 20* (pp. 219-233). Springer International Publishing.
- [40] Alam, M.S., & Vuong, S. T. (2013, August). Random forest classification for detecting android malware. In 2013 IEEE international conference on green computing and communications and IEEE Internet of Things and IEEE cyber, physical and social computing (pp. 663-669). IEEE.
- [41] Farnaaz, N., & Jabbar, M. A. (2016). Random forest modeling for network intrusion detection system. *Procedia Computer Science*, 89, 213-217.
- [42] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1), 3-24.
- [43] Oisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3), 128-138.

Collateral Circulation Classification Based on Cone Beam Computed Tomography Images using ResNet18 Convolutional Neural Network

Nur Hasanah Ali¹, Abdul Rahim Abdullah², Norhashimah Mohd Saad³, Ahmad Sobri Muda⁴

Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia¹

Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia²

Faculty of Electrical & Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia³

Department of Imaging-Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia⁴

Abstract—Collateral circulation is an arterial anastomotic channel that supply nutrient perfusion to areas of the brain. It happens when there is an existence of disruption of regular sources of flow due to an ischemic stroke. The most recent method, Cone Beam Computed Tomography (CBCT) neuroimaging is able to provide specific details regarding the extent and adequacy of collaterals. The current approaches for collateral circulation classification are based on manual observation and lead to inter and intra-rater inconsistency. This paper presented a 2-class automatic classification that is recently growing very fast in artificial intelligence disciplines. The two classes will differentiate between good and poor collateral circulation. A pre-trained convolutional neural network (CNN), namely ResNet18, has been used to learn features and train using 4368 CBCT images. Initially, the dataset is prepared, labeled and augmented. Then the images were transferred to be trained using the ResNet18 method with certain specifications. The algorithm performance was then evaluated using metrics in terms of accuracy, sensitivity, specificity, F1 score and precision on the CBCT images to classify collateral circulation accurately. The findings can automate collateral circulation classification to ease the limitations of standard clinical practice. It is a convincing method that supports neuroradiologists in assessing clinical scans and helps neuroradiologists in clinical decisions about stroke treatment.

Keywords—Collateral circulation; CBCT; ResNet; convolutional neural network; classification

I. INTRODUCTION

Stroke disease is one of the causes that lead to short or long-term disability in developed countries. Stroke disease is also one of the top causes of mortality in the world [1]. Worldwide, over 5.5 million annual mortality rate has been reported, while 50% became disabled as a result of their strokes [2]. Women had poorer post-stroke outcomes and were more likely to experience a stroke in their lifetime [3]. In 2019, the low-income group had a higher age-standardized stroke-related death rate than the high-income group [4]. Most strokes are often caused by the obstruction of pathways by both the brain and heart. The impact of stroke can be minimized by early detection of warning signs [5],[6]. Stroke disease is divided into two categories or groups: hemorrhagic stroke and ischemic stroke [6]. Most ischemic strokes will occur due to an unpredicted obstruction in the blood flow to several areas of

the brain. Lack of oxygen and nutrients for the cells in those areas of the brain will cause the cells death [5] and lead to other serious problems such as blood vessel ruptures, also known as a hemorrhagic stroke when the brain tissue is bleeding [7]. Although thrombectomy carries inherent risks, it should only be performed in stroke disease patients with certain signs, which are a large penumbra and small infarct, along with collateral circulation [1,2].

In the case of acute brain ischemia, cerebral collateral circulation plays a vital role in compensatory mechanisms [8]. As a result of a failure of the primary arteries, the cerebral collateral circulatory system acts as a secondary network of vessels pathway that maintains cerebral blood flow [9]. Good collateral circulation and a lower likelihood of hemorrhagic transformation should improve endovascular treatment for acute ischemic stroke [10]. Extending the therapeutic time window after ischemia and boosting collateral blood flow perfusion are essential components of treating ischemic stroke [6]. It has been shown that good collateral circulation makes a significant difference in the functional outcome [11] and recurrence risk of stroke patients suffering from different causes and receiving medical or endovascular treatment. Several features have been investigated to diagnose the conditions of collateral circulation and compare findings with stroke disease patients. Assessment of ischemic stroke of collateral circulation is actively investigated. As collateral circulation is critical in the assessment of penumbra presence and volume, which are critical factors in the severity and time course of ischemic strokes, the status of collateral circulation is critical [11], [12]. Fig. 1 shows the collateral circulation view in the human brain. However, rather than measuring the actual anatomical connections, these approaches assess the general condition of collaterals.

Imaging modality technique using Magnetic Resonance Imaging (MRI), Computerized Tomography (CT) [13], X-ray, CBCT, etc., provides precise details regarding the flow of blood to the various parts of the brain [14]. Then, when the imaging surveys have been completed, a comprehensive neurological examination must be undertaken [15]. These characteristics determine whether the underlying brain parenchyma survives in comparison to an arterial lesion. Cone Beam Computed Tomography (CBCT) is one of the most

popular techniques for assessing many diseases, especially the collateral circulation in the brain [16]. CBCT is considered an advanced imaging technology that provides accurate and three-dimensional (3D) images for assessing hard tissue, soft tissue, and bone [17],[18]. As a result of its advantages over conventional CT, CBCT is increasingly used in acute strokes and neurovascular image-guided procedures [19], including strokes and nerve damage. Fig. 2 shows an example of CBCT images.

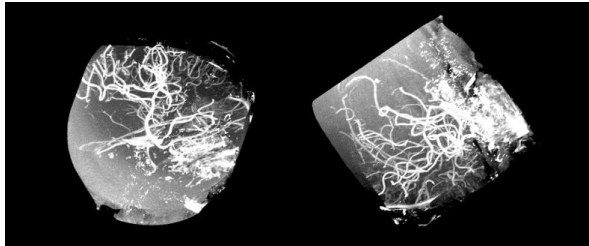


Fig. 1. Collateral circulation in the human brain.

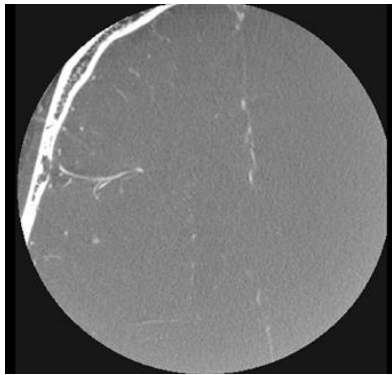


Fig. 2. CBCT image.

In recent years, machine learning specifically deep learning has become increasingly popular. Deep learning is a type of learning technique that employs multi-layered neural networks [15]. It has shown promising results in retrieving useful information from medical images and signals [20]. This research demonstrated an analysis framework to classify collateral circulation accurately for ischemic stroke patients into two classes: good and poor. The proposed method has been chosen to capture complex, non-homogeneous structures and tiny-size images. The aim is to discover the utilization of deep learning techniques to automate the classification of collateral circulation on CBCT images.

II. RELATED WORK

A. Collateral Circulation Scoring

Collateral circulation is an alternative network vessels that carries blood to the same destination tissue [21]. It serves as an auxiliary vascular system and plays a crucial role in preventing cerebral ischemia when the primary vascular pathways are partially obstructed [22]. Table I presents the state-of-the-art evidence suggesting that the combination of neuroradiology expertise and artificial intelligence holds promise in facilitating timely and accurate disease diagnosis.

TABLE I. COLLATERAL CIRCULATION GRADING SYSTEMS

Author	Modality	Grading System
Kucinski et al. [23]	Cerebral angiography	1 (good): ≥ 3 MCA branches (retrograde filling) 2 (poor): < 3 MCA branches
Higashida et al. [24]	Cerebral angiography	0: no collateral vessels filled 1: slow collateral filling to periphery 2: rapid collateral filling to periphery 3: collaterals with slow but complete flow in ischaemic bed 4: rapid and complete flow in entire ischaemic territory
Maas et al. [25]	CT angiography	1: absent 2: less than contralateral side 3: equal to contralateral side 4: greater than contralateral side 5: exuberant
Silvestrini et al. [26]	Transcranial doppler	Collateral supply inferred by direction of flow in ophthalmic artery, anterior cerebral artery, and posterior cerebral artery 1: Good: ≥ 2 vessels insonated 2: Poor: ≤ 1 vessel insonated
Miteff et al. [27]	CT angiography	1 (good): entire MCA distal to occlusion reconstituted with contrast 2 (moderate): some branches of MCA reconstituted in Sylvian fissure 3 (poor): distal superficial branches reconstituted
Tan et al. [28]	CT angiography	0: absent 1: $< 50\%$ collateral MCA filling 2: $> 51-99\%$ 3: 100%
Lee et al. [29]	MRI, magnetic resonance angiography	Distal hyperintense vessels on FLAIR MRI 1: absent 2: subtle 3: prominent
Marta. [30]	CT angiography	1: Good (100% collateral supply of the occluded MCA territory); 2: Intermediate (collateral supply filling $> 50\%$ but $< 100\%$ of the occluded MCA territory) or 3: Poor (collateral supply filling $\leq 50\%$ but $> 0\%$ of the occluded MCA territory)
Jiahang Su [31]	CT angiography	0: absent collaterals (0% filling in occluded territory) 1: poor collaterals ($> 0\%$ and 50% filling in occluded territory) 2: moderate collaterals ($> 50\%$ and $< 100\%$ in occluded territory) 3: good collaterals (100% filling in occluded territory)
Proposed method	CBCT	1: good collaterals (collateral supply $> 50\%$ and $< 100\%$) 2: poor collaterals (collateral supply $> 0\%$ and 50%)

Su et al. and Tan et al. proposed a four-grade scoring system to prove a correlation between the outcome and effect of Endovascular Thrombectomy (EVT). Silvestrini et al. studied 66 patients having cervical arterial dissection. The researchers showcased the potential of Transcranial Doppler (TCD), a non-invasive technique, in assessing the long-term prognosis of patients in such cases. TCD was employed within 24 hours of a stroke associated with carotid dissection to evaluate the collateral status.

Maas et al. and Higashida et al. rated using a five-point scale for collateral circulation viewed during CT angiography. In the study conducted by Maas et al., a reference group of 235 patients without occlusions was included, along with 134 patients with acute stroke and MCA occlusion. The study aimed to assess the severity of ischemic stroke, prehospital clinical fluctuations, and clinical deterioration in the days following hospital admission. Additionally, the impact of collaterals was also evaluated in the study. After 1 hour of the onset of symptoms, poor collaterals were visible in 38% of patients; this number fell to 12% in patients whose images were taken 12 to 24 hours later. Patients with inadequate collaterals did not experience any variations in prehospital symptoms. Those with insufficient collaterals, as opposed to those with normal or voluminous collaterals, had a four times higher likelihood of experiencing symptom deterioration while hospitalized.

Miteff et al. and Kersten-Oertel et al. used three grading systems. Kersten-Oertel et al. developed a technique for variations of mean intensities between the left and right hemispheres. The computed score and the neuroradiologist's assessment correlated well ($r^2 = 0.71$), but the approach itself had difficulty for individual variations, such as those resulting from calcification and normal vasculature asymmetry between hemispheres. Miteff et al. employed a grading system consisting of three levels to assess the collateral circulation. A grade of three was assigned when the vessels were observed to be reconstituted beyond the occlusion site. A grade of 2 indicated the presence of visible vessels at the Sylvian fissure. A grade of one denoted the situation where contrast opacification was only observed in the distal superficial branches. In their study, 55% of the patients had good collaterals, 26% had moderate collaterals, and 18% had poor collaterals.

B. Deep Learning in Ischemic Stroke Analysis

There are several works already published to automate diagnosis decisions in ischemic stroke classification. Raj et al. introduced a novel approach that combined ResNet50 and ViT in their study. The combined model achieved an accuracy of 87%. When evaluating the detection of hemorrhage, infarct, and normal cases, the true positive rates were 0.77, 0.76, and 0.91, respectively. The study involved a total of 233 patients, out of which 70 had infarcts, 67 had hemorrhages, and 96 were classified as normal. It is worth noting that the number of slices depicting hemorrhage and infarct was relatively low, as these conditions typically occur in specific brain areas that are visible in only a limited number of CT scan slices. In their

study, Wei et al. introduced a novel classification approach called Semantic Segmentation Guided Detector Network (SGD-Net). The technique combines DenseUNet121, ResUNet50, and VGGUNet16 models for the classification of DWI images in 216 acute ischemic stroke patients. The DWI images had a scale of 384×384 pixels per transverse slice, with each patient having 20 to 28 serial transverse slices.

Gautam and Raman conducted a comparison of their technique with other CNN models, including AlexNet, ResNet50, P_CNN_WP, and P_CNN. The authors introduced a framework specifically designed for the classification of brain CT images into hemorrhagic, ischemic, and normal categories using 2D CT scan slice images. Rajendran et al. conducted three experiments to classify CT slices of ischemic stroke patients. The third approach using an ensemble model (ResNet50, VGG16, and InceptionV3) achieved an accuracy of 81.98%. Ozaltine et al. used OzNet method combined with other method such as minimum Redundancy Maximum Relevance (mRMR) method and Decision Tree (DT), k-Nearest Neighbors (kNN), Linear Discriminant Analysis (LDA), Naïve Bayes (NB), and Support Vector Machines (SVM) to achieve high classification performance. As a result, the new method OzNet-mRMR-NB is able to classify strokes with an accuracy of 98.42%. Eshmawi et al. developed a binary classification using new CAD-BSDC model for MRI images. The simulation results showed that the proposed CAD-BSDC technique was more effective than the most recent state-of-the-art approaches in terms of a variety of performance measures.

Recently, study by Sercan et al. examined the deep learning method for stroke classification. The U-Net, a method proposed in this study, utilizes encoder-decoder architecture. This architecture, which is based on deep learning, is highly effective in addressing various challenges in artificial intelligence applications. The results of the study indicate exceptional performance of the proposed model, with accuracy rates of 98.9% for stroke classification and 98.5% for ischemia and hemorrhage classification. Govindarajan et al. gathered data on 507 patients as part of a study by classifying stroke disorders using a text mining combination and a machine learning classifier. They employed ANN to train multiple machine learning techniques for their analysis, and the SGD method provided them with the best value, which was 95%.

In this study, a deep transfer residual convolution neural network structure named ResNet18 is proposed to classify collateral circulation using CBCT images. This method was selected due to ease in residual mapping and shortcut connections lead to better results compared to very deep plain networks [32]. In addition, using the ResNet method, the training process is easier and the performance is sustained even though the architecture is getting deeper [32]–[34]. Thus, this proposed method is able to help neuroradiologists to speed up the treatment decision

III. METHODOLOGY

The classification proposed method can be described using the flowchart in Fig. 3.

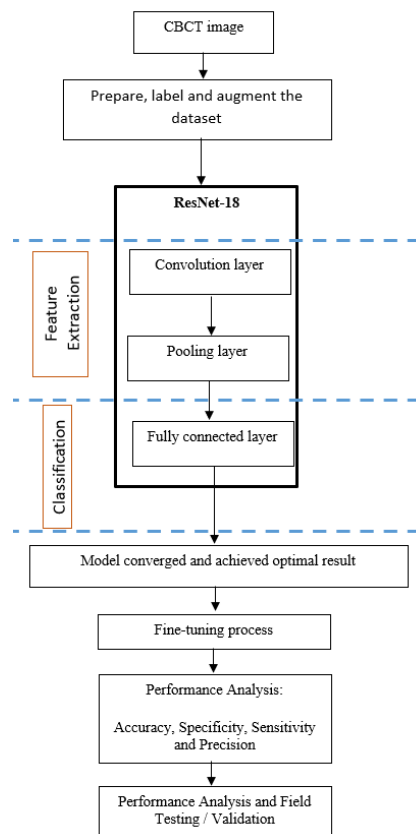


Fig. 3. Research flow for proposed method.

C. Materials

For this study, we included 30 patients who had suffered an ischemic stroke. For all subjects, CBCT imaging was acquired on a Philips VasoCT scanner. The VasoCT upgrades are supported on the Philips Allura Xper systems provided with XperCT. The VasoCT acquisitions are performed with a motorized rotational C-arm movement and result in an isotropic stack of VasoCT images that can be visualized in any random position without image quality loss. All samples have medical records which have been confirmed by neuroradiologists. Images were encoded in DICOM (Digital Imaging and Communications in Medicine) format. The research mainly focuses on the process of classification by using CNN techniques using Python as the computational tool. This research does not include clinical representation, patient history, historical findings, or present solutions for the lesion.

Based on the collected data, automatic classification is implemented using ResNet18 models. The research mainly focuses on the process of classification by using ResNet18 models using Python as the computational tool. The deep learning framework is PyTorch. The Jupyter Notebook compiler that belongs to the Anaconda package was used in addition to some other basic Python libraries such as Numpy, Pillow, Augmentor, and OpenCV.

D. Deep Learning Model using ResNet18

ResNet networks have been developed based on the concept of residual learning [35][36][37]. This technique is one of the popular techniques in the deep learning model developed

by He et al. in 2016 [32]. Residual learning is the learning process that involves a residual connection [33]. Residual connections are the connections that link the output of previous layers to the output of new layers [38]. A residual neural network (ResNet) is a supervised learning algorithm that is based on prismatic cell constructions in the cerebral cortex. Individual things or bypasses are utilized by ResNet18 to hop past certain levels. The most common residue neural network models include double or triple-layer delays [39] with nonlinearities (ReLU) and average pooling in between. To train the bypass values, an extra weight vector can be utilized; those models can be categorized as HighwayNets. DenseNets are networks that have multiple simultaneous bypasses. A non-residual network can be defined as a straightforward system in the setting of Convolution Neural Network models [34],[40], [41].

There are two major reasons to use hidden layers: to prevent diminishing slopes and to alleviate the Depreciation (precision overload) phenomenon, which occurs when adding additional layers to relatively deep network results in increased generalization error. The weights adjust throughout learning to muffle the previous layer and magnify the recently bypassed element. Only the values for the neighboring element's link are changed inside the basic instance[42], with no specific values for the downstream layers. While a unique nonlinear layer is passed over, or when the middle layers are all normal, this approach has good performance.

The functionality of ResNet18 for collateral circulation classification has been investigated in this research. The model depth is represented by the number "18." From the first to the deep network, the system complexity is defined as the highest number of successive convolution operations and fully linked layers on a path. The ResNet18 models that were used are given along with their specifications. ResNet18 algorithms are suitable for two-dimensional and three-dimensional methods, with the dimensions of filtration systems and source images (which might be two-dimension or three-dimension) differing. To suit CBCT scans, the updated 3D ResNet18 utilizes lesser data and has stride '1' in the first convolution operation.

ResNet18 has a good performance to another model of ResNets, but because it is deep, it may reduce characteristics. As a result, we employ the ResNet18 pre-trained model as a feature representation (encoder) for our network structure. ResNet18 has 16 convolutional layers and several fully connected layers (Fc). The input image of ResNet is 224x224, the pooling operation is 77 pixels, and the remaining layers are 33 pixels. After average pooling, the fully connected convolution layer extracted features, and the network yields a wavelet coefficient, which is then processed with Softmax to get the categorization rate. There is the same amount of layers in the convolution layer that produces the same size extracted features. ResNet18 will produce a wavelet coefficient with several values, which are used to signal that the input picture corresponds to a specific category, and the outcome will be the class with the greatest chance. Because the fully connected FC keyframe input connections must be limited [43], ResNet18's raw image must be adjusted in size.

Based on this concept of residual connection as shown in Fig. 4, researchers could develop more than one architecture such as ResNet18, ResNet34, ResNet50, ResNet100, and Inception-ResNet, all of which have shown very high accuracy in comparison to those networks that do not have residual connections. To imagine what ResNet18 looks like, imagine 18 weighted layers all interconnected with residual connections. It starts with a convolutional layer that has 64 filters, a kernel size of 7x7, and a stride of two then it goes through a pooling layer that has two strides, and so on till the information reaches the fully connected layer. The dotted shortcuts indicate an increase in dimensions to be able to concatenate with the next layer [44].

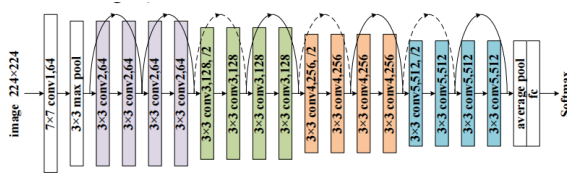


Fig. 4. Original ResNet18 architecture.

E. Performance Evaluation

The performance evaluation of the ResNet18 model included measures such as accuracy, sensitivity, specificity, precision, and F1 score. These are the numerical measurements of the model's performance, where accuracy is defined as the proportion of accurately detected samples to the total number of samples. Specificity and sensitivity are measurements of correctly identifying two different classes, which are, by definition, negative and positive.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ number\ of\ samples} \quad (1)$$

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (2)$$

$$Sensitivity = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (4)$$

$$F1\ score = \frac{2 \times Precision \times Sensitivity}{Precision + Sensitivity} \quad (5)$$

IV. RESULTS AND DISCUSSION

Different collateral circulation classification was performed on the above-mentioned dataset using the ResNet18 model. We divide the training data and validation data by 80:20, which means 80% training data and 20% validation data. In this ResNet18 classification technique, based on the seven epochs, the result has achieved the best accuracy of 0.6590, as shown in Table II.

TABLE II. ACCURACY FOR RESNET18 METHOD

Epoch	Testing
1	0.659
2	0.548
3	0.613
4	0.557
5	0.601
6	0.578
7	0.534

To the best of my knowledge, no previous research of a similar nature has been conducted due to the limitations inherent in this study. While there is existing research focused on classifying CTA and MRI images [45]–[47], the investigation into collateral circulation based on CBCT images using deep learning techniques is relatively novel and scarce. This study contributes to bridging this gap in the literature by exploring the potential of CBCT images and deep learning algorithms for collateral circulation classification.

In this research, the performance of a model is also evaluated by using the training and testing loss measures respectively, during the training and testing phases of the process. The model is trained on a set of input data while in the training phase, and the training loss is calculated after each iteration of the training process. The training loss measures how well the model can forecast the output based on the information provided in the input. During the training phase, the goal is to achieve the best possible results with the least amount of loss. This is often accomplished by modifying the model's weights and biases by applying an optimization procedure such as stochastic gradient descent.

Fig. 5 presents the training and testing loss graph, which provides valuable insights into the performance of the model. The graph indicates that the loss during the training phase remains relatively low, indicating that the model is learning effectively from the training data. However, a notable observation is that the loss during the testing phase is significantly higher than the training loss, suggesting the presence of overfitting. Overfitting occurs when a model becomes too specialized to the training data and struggles to generalize well to unseen data. To address this issue, several modifications can be implemented. One effective approach is to introduce regularization techniques. Another strategy to combat overfitting is to increase the size of the dataset. By obtaining more diverse and representative data, the model can learn from a wider range of examples and become more resilient to overfitting.



Fig. 5. Training and testing loss comparison graph.

The performance evaluation metrics can be calculated using Eq. (1) and (5). It is calculated that the accuracy is 0.660, sensitivity is 0.776, specificity is 0.526, precision is 0.650 and F1 score is 0.698. The sensitivity rate of the experiments shows that the CBCT scan was detected as positive for collateral circulation. The high sensitivity of the suggested model can

offer neuroradiologists a ‘second opinion’. The dataset is assessed using the confusion matrix obtained from the experiment as shown in Fig 6. The confusion matrix provides in-depth explanations of the model's test outcomes. The confusion matrix provides a thorough examination of the correct and wrong classifications for this model class. Additionally, the confusion matrices demonstrate that some samples are incorrectly classified; indicating that the model is confused and unable to determine which class is the correct class for the incorrectly classified sample. This research can aid in providing a quick and precise diagnosis when compared to experimental tests, which require more time and have a higher likelihood of producing false negative results.

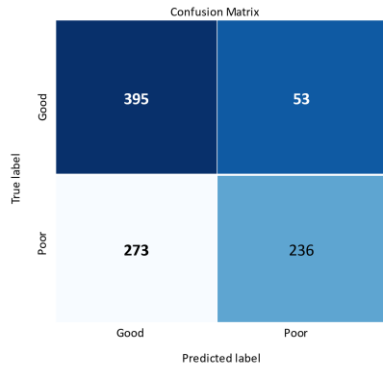


Fig. 6. Confusion matrix for testing data.

V. CONCLUSION

In conclusion, a novel fully automatic approach to classifying the different stages of collateral circulation from CBCT images using the ResNet18 model of CNN is proposed. The data augmentation technique was used to increase the total number of 4368 CBCT images for training and testing for seven training epochs. Two stages of collateral circulation, good and poor were classified. The best result of 65.9 % accuracy was obtained. With this technique, it will be easier to detect collateral circulation classes fast and its treatment procedure will be more comprehensive. Despite the achievements reported in this paper, several improvements remain possible. Future research in the domain shall address these issues, possibly with a higher number of data to get a better training effect and further tuning of the transfer learning model.

ACKNOWLEDGMENT

The authors would like to thank the Multimedia University (MMU) that supported this research, and all involved from the Advanced Digital Signal Processing (ADSP) Research Lab, Centre of Robotic and Industrial Automation (CeRIA), Faculty of Electrical Engineering (FKE), Faculty of Electrical and Electronic Engineering Technology (FTKEE), Universiti Teknikal Malaysia Melaka (UTeM) and Department of Imaging, Hospital Pengajar Universiti Putra Malaysia (HPUPM).

REFERENCES

[1] M. S. Phipps and C. A. Cronin, "Management of acute ischemic stroke," *BMJ*, vol. 368, 2020.

[2] E. S. Donkor, "Stroke in the 21st Century: A Snapshot of the Burden, Epidemiology, and Quality of Life," *Stroke Res. Treat.*, vol. 2018, 2018.

[3] C. W. Yoon and D. Bushnell, "Stroke in Women : A Review Focused on Epidemiology , Risk Factors , and Outcomes," *J. Stroke*, vol. 25, no. 1, pp. 2–15, 2023.

[4] V. L. Feigin et al., "Global, regional, and national burden of stroke and its risk factors, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019," *Lancet Neurol.*, vol. 20, no. 10, pp. 1–26, 2021.

[5] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance Analysis of Machine Learning Approaches in Stroke Prediction," *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, pp. 1464–1469, 2020.

[6] B. C. V. Campbell et al., "Ischaemic stroke," *Nat. Rev. Dis. Prim.*, vol. 5, no. 1, 2019.

[7] S. Paul and E. Candelario-Jalil, *Emerging neuroprotective strategies for the treatment of ischemic stroke: An overview of clinical and preclinical studies*, vol. 335. Elsevier Inc, 2021.

[8] K. Malhotra and D. S. Liebeskind, "Collaterals in ischemic stroke," *Brain Hemorrhages*, vol. 1, no. 1, pp. 6–12, 2020.

[9] H. E. Vasquez et al., "Intracranial collateral circulation and its role in neurovascular pathology," *Egypt. J. Neurosurg.*, vol. 36, no. 1, pp. 0–4, 2021.

[10] O. Y. Bang et al., "Collateral flow averts hemorrhagic transformation after endovascular therapy for acute ischemic stroke," *Stroke*, vol. 42, no. 8, pp. 2235–2239, 2011.

[11] L. Liu et al., "Guidelines for evaluation and management of cerebral collateral circulation in ischaemic stroke 2017," *Stroke Vasc. Neurol.*, vol. 3, no. 3, pp. 117–130, 2018.

[12] G. S. Piedade et al., "Cerebral Collateral Circulation: A Review in the Context of Ischemic Stroke and Mechanical Thrombectomy," *World Neurosurg.*, vol. 122, pp. 33–42, 2019.

[13] N. H. Ali, A. R. Abdullah, N. Mohd Saad, A. S. Muda, T. Sutikno, and M. H. Jopri, "Brain stroke computed tomography images analysis using image processing: A Review," *IAES Int. J. Artif. Intell.*, vol. 10, no. 4, p. 1048, 2021.

[14] C. M. Lo, P. H. Hung, and D. T. Lin, "Rapid Assessment of Acute Ischemic Stroke by Computed Tomography Using Deep Convolutional Neural Networks," *J. Digit. Imaging*, vol. 34, no. 3, pp. 637–646, 2021.

[15] and A. B. H. Anas Tharek, Ahmad Sobri Muda, Aqilah Baseri Hudi, "INTRACRANIAL HEMORRHAGE DETECTION IN CT SCAN USING DEEP LEARNING," *Asian J. Med. Technol. Vol. 2 No. 1*, vol. 2, no. 1, pp. 1–18, 2022.

[16] A. Abd Aziz et al., "Detection of Collaterals from Cone-Beam CT Images in Stroke," *Sensors*, vol. 21, no. 23, p. 8099, 2021.

[17] S. Lata, S. K. Mohanty, S. Vinay, A. C. Das, S. Das, and P. Choudhury, "Is Cone Beam Computed Tomography (CBCT) a Potential Imaging Tool in ENT Practice?: A Cross-Sectional Survey Among ENT Surgeons in the State of Odisha, India," *Indian J. Otolaryngol. Head Neck Surg.*, vol. 70, no. 1, pp. 130–136, 2018.

[18] K. J. Jeon, C. Lee, Y. J. Choi, and S. S. Han, "Comparison of the usefulness of CBCT and MRI in TMD patients according to clinical symptoms and age," *Appl. Sci.*, vol. 10, no. 10, 2020.

[19] P. Nicholson et al., "Novel flat-panel cone-beam CT compared to multi-detector CT for assessment of acute ischemic stroke: A prospective study," *Eur. J. Radiol.*, vol. 138, p. 109645, 2021.

[20] J. Too, A. R. Abdullah, N. M. Saad, and W. Tee, "EMG feature selection and classification using a Pbest-guide binary particle swarm optimization," *Computation*, vol. 7, no. 1, 2019.

[21] A. Sharma, A. Agarwal, V. Y. Vishnu, and M. V. P. Srivastava, "Collateral Circulation- Evolving from Time Window to Tissue Window," *Ann. Indian Acad. Neurol.*, vol. 26, no. 1, pp. 10–16, 2022.

[22] T. Verdolotti et al., "Colorviz, a new and rapid tool for assessing collateral circulation during stroke," *Brain Sci.*, vol. 10, no. 11, pp. 1–8, 2020.

[23] T. Kucinski et al., "Collateral circulation is an independent radiological predictor of outcome after thrombolysis in acute ischaemic stroke," *Neuroradiology*, vol. 45, no. 1, pp. 11–18, 2003.

- [24] R. T. Higashida et al., "Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke.," *Stroke.*, vol. 34, no. 8, 2003.
- [25] M. B. Maas et al., "Collateral vessels on CT angiography predict outcome in acute ischemic stroke.," *Stroke.*, vol. 40, no. 9, pp. 3001–3005, 2009.
- [26] M. Silvestrini et al., "Early activation of intracranial collateral vessels influences the outcome of spontaneous internal carotid artery dissection," *Stroke*, vol. 42, no. 1, pp. 139–143, 2011.
- [27] F. Miteff, C. R. Levi, G. A. Bateman, N. Spratt, P. McElduff, and M. W. Parsons, "The independent predictive utility of computed tomography angiographic collateral status in acute ischaemic stroke," *Brain*, vol. 132, no. 8, pp. 2231–2238, 2009.
- [28] I. Y. L. Tan et al., "CT angiography clot burden score and collateral score: Correlation with clinical and radiologic outcomes in acute middle cerebral artery infarct," *Am. J. Neuroradiol.*, vol. 30, no. 3, pp. 525–531, 2009.
- [29] K. Y. Lee, L. L. Latour, M. Luby, A. W. Hsia, J. G. Merino, and M. P. S. Warach, "Distal hyperintense vessels on FLAIR: An MRI marker for collateral circulation in acute stroke?," *Neurology*, vol. 72, no. 13, pp. 1134–1139, 2009.
- [30] M. Kersten-Oertel, A. Alamer, V. Fonov, B. W. Y. Lo, D. Tampieri, and D. L. Collins, "Towards a computed collateral circulation score in ischemic stroke," *Comput. Vis. Intravasc. Imaging Comput. Assist. Stenting (CVII STENT)*, no. August 2017, 2017.
- [31] J. Su et al., "Automatic Collateral Scoring From 3D CTA Images," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 2190–2200, 2020.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016.
- [33] D. Sarwinda, R. H. Paradisa, A. Bustamam, and P. Anggia, "Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer," *Procedia Comput. Sci.*, vol. 179, no. 2019, pp. 423–431, 2021.
- [34] M. Gao, P. Song, F. Wang, J. Liu, A. Mandelis, and D. Qi, "A Novel Deep Convolutional Neural Network Based on ResNet-18 and Transfer Learning for Detection of Wood Knot Defects," *J. Sensors*, vol. 2021, 2021.
- [35] H. Hu et al., "Content-based gastric image retrieval using convolutional neural networks," *Int. J. Imaging Syst. Technol.*, vol. 31, no. 1, pp. 439–449, 2021.
- [36] S. Ma, T. Huang, X. Sun, and Y. Wei, "Driver Drowsiness Detection Based on ResNet-18 and Transfer Learning," *Proc. 33rd Chinese Control Decis. Conf. CCDC 2021*, pp. 2390–2394, 2021.
- [37] P. V. Rouast, M. T. P. Adam, and R. Chiong, "Deep Learning for Human Affect Recognition: Insights and New Developments," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 524–543, 2021.
- [38] A. Ebrahimi, S. Luo, and R. Chiong, "Introducing Transfer Learning to 3D ResNet-18 for Alzheimer's Disease Detection on MRI Images," *Int. Conf. Image Vis. Comput. New Zeal.*, vol. 2020-Novem, 2020.
- [39] E. Jing, H. Zhang, Z. G. Li, Y. Liu, Z. Ji, and I. Ganchev, "ECG Heartbeat Classification Based on an Improved ResNet-18 Model," *Comput. Math. Methods Med.*, vol. 2021, 2021.
- [40] S. Liu, G. Tian, and Y. Xu, "A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter," *Neurocomputing*, vol. 338, pp. 191–206, 2019.
- [41] X. Yu and S. H. Wang, "Abnormality Diagnosis in Mammograms by Transfer Learning Based on ResNet18," *Fundam. Informaticae*, vol. 168, no. 2–4, pp. 219–230, 2019.
- [42] T. Shaily and S. Kala, "Bacterial Image Classification Using Convolutional Neural Networks," *2020 IEEE 17th India Counc. Int. Conf. INDICON 2020*, 2020.
- [43] P. Ghosal, L. Nandanwar, S. Kanchan, A. Bhadra, J. Chakraborty, and D. Nandi, "Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network," *2019 2nd Int. Conf. Adv. Comput. Commun. Paradig. ICACCP 2019*, pp. 1–6, 2019.
- [44] M. Guo and Y. Du, "Classification of Thyroid Ultrasound Standard Plane Images using ResNet-18 Networks," *Proc. Int. Conf. Anti-Counterfeiting, Secur. Identification, ASID*, vol. 2019-Octob, no. i, pp. 324–328, 2019.
- [45] G. Tetteh et al., "A deep learning approach to predict collateral flow in stroke patients using radiomic features from perfusion images," *Front. Neurol.*, vol. 14, no. 1, 2023.
- [46] H. Kuang, W. Wan, Y. Wang, J. Wang, and W. Qiu, "Automated Collateral Scoring on CT Angiography of Patients with Acute Ischemic Stroke Using Hybrid CNN and Transformer Network," *Biomedicines*, vol. 11, no. 2, 2023.
- [47] L. Hokkinen, T. Mäkelä, S. Savolainen, and M. Kangasniemi, "Computed tomography angiography-based deep learning method for treatment selection and infarct volume prediction in anterior cerebral circulation large vessel occlusion," *Acta Radiol. Open*, vol. 10, no. 11, p. 205846012110603, 2021.

An Enhanced Algorithm of Improved Response Time of ITS-G5 Protocol

Kawtar Jellid¹, Tomader Mazri²

Dept. of Electronic System, Information Processing, National School of Applied Sciences, Kenitra, Morocco¹

Abstract—This research article proposes an algorithm for improving the ITS-G5 protocol, which addresses the issue of response time. The algorithm includes the integration of Dijkstra's algorithm to prioritize shorter paths for message transmission, resulting in reduced delays. The initial algorithm for the ITS-G5 protocol is presented, followed by the modified algorithm that incorporates Dijkstra's algorithm. The modified algorithm utilizes a node-based approach and implements Dijkstra's algorithm to find the shortest path between two nodes. The algorithm is evaluated in a scenario involving 20 vehicles, where each vehicle has its own message. The results show improved communication efficiency and reduced response time compared to the original ITS-G5 protocol.

Keywords—ITS-G5 (Intelligent Transport Systems); V2V (Vehicle-to-Vehicle); V2I (Vehicle-to-Infrastructure); V2X (Vehicle-to-everything); autonomous vehicle

I. INTRODUCTION

Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, also known as vehicle-to-everything (V2X), involve wireless technology with the aim of facilitating data exchanges between an automobile and its environment. In recent times, two primary standards for vehicular communication have emerged, utilizing the specially allocated 5.9GHz unlicensed frequency band. On one hand, there is the Dedicated Short-Range Communications (DSRC) protocol, developed in the United States [1]. On the other hand, there is the Intelligent Transportation System (ITS-G5) protocol, created by the European Telecommunications Standards Institute (ETSI).

These standards are built upon the IEEE 802.11p access layer, which was specifically designed for communication within vehicular networks, [2]. Additionally, the connectivity layer specified in the European "Delegated Act" is built upon the IEEE 802.11p standard for vehicular networks [3]. The objective of intelligent transportation systems (ITS) is to enhance traffic security, effectiveness, and the convenience of automobile occupants by utilizing different detectors, gadgets, physical structures, and communication technologies.

Collaborative-ITS systems facilitate direct links between automobiles V2V communication or between automobiles and infrastructure (V2I or I2V communications). These links are supported by onboard units, services, and specialized devices that employ specific interfaces between automobiles, susceptible road users (VRUs), and roadside units (RSUs) [4]. To summarize, the Intelligent Transportation System (ITS) serves as a catalyst for enhanced road safety and the advancement of autonomous vehicle technologies.

Additionally, it aims to improve traffic efficiency by promoting smoother and more efficient flow of vehicles. The scope of ITS encompasses various applications, including driver convenience, public transportation, and commercial transportation of goods.

Within this framework [5], the concept of vehicle-to-everything (V2X) communication refers to real-time communication within the transportation domain.

This communication paradigm facilitates seamless interactions and data exchange between vehicles, infrastructure, pedestrians, and other entities, enabling the realization of innovative and interconnected transportation solutions.

ITS-G5 quickly established itself as a catalyst, spurring the rapid development of state-of-the-art applications in the field of traffic efficiency and safety [6]. Its implementation as a reliable communication framework for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) interactions paved the way for significant advancements and breakthroughs in this domain. ITS-G5 is an access technology developed by ETSI specifically designed to enable communication between vehicles, infrastructure, and various ITS services.

This communication is facilitated through the encapsulation of safety and non-safety applications into Cooperative Awareness Message (CAM) [7] and Decentralized Environmental Notifications Message (DENM).

These messages are then encapsulated into Geo-networking messages and transmitted using the Basic Transport Protocol (BTP) at the access layer, while being regulated by the Decentralized Congestion Control (DCC) mechanism.

Like several protocols, the ITS-G5 protocol algorithm has drawbacks such as interoperability, spectrum availability, and deployment challenges. However, one of the major concerns is the response time, which significantly affects the efficiency of communication between vehicles.

That is why the main objective of this paper is to propose an improved algorithm for the ITS-G5 protocol, aiming to reduce and optimize the response time by incorporating the Dijkstra algorithm, by utilizing the Dijkstra algorithm, the ITS-G5 protocol can give priority to the most optimal routes for message transmission, thereby reducing time lags and enhancing the overall effectiveness of communication. This empowers vehicles to make quicker and well-informed judgments grounded on real-time information, amplifying their capacity to promptly react to significant occurrences or potential dangers on the road.

II. RELATED WORK

Multiple investigations, primarily through simulation-based studies, have been carried out to examine the performance of ITS-G5 protocols. However, the performance of the ITS-G5 protocol falls short, particularly in the context of platooning. A comparison of results with and without data traffic from regular vehicles reveals that greater reliability is achieved when there is no additional data traffic from regular vehicles [8], and when the Cooperative Control Channel (CCH) is exclusively dedicated to interpolation communication. This noteworthy enhancement in performance can be attributed to a significant reduction in packet collisions.

BRISA, a prominent Portuguese highway and mobility services provider, engaged in a collaborative effort with the Institute of Telecommunications (IT) [9] to address the complex challenges inherent in intelligent vehicular networks (IVNs) these challenges posed by latency and throughput.

The primary focus of their joint endeavor was to tackle the issues of latency and throughput, particularly within the context of the emerging IEEE 802.11p standard. These challenges arise due to the dynamic nature of vehicular networks, where variable vehicle speeds disrupt connectivity and necessitate frequent recalculations for effective node coordination. In response, BRISA and IT embarked on a series of research initiatives designed to enhance communication performance within the demanding conditions of IVNs. Their collaborative efforts encompassed the development of novel technologies, the establishment of real-world experimentation platforms, rigorous testing and validation processes, and the formulation of advanced communication protocols. Through these concerted efforts, the goal was to optimize latency, throughput, and communication reliability, thereby contributing to the advancement of safer, more efficient intelligent transportation systems.

Scientific research has provided evidence that the response time of the ITS-G5 protocol is indeed a significant drawback. Numerous studies have consistently shown that the protocol's prolonged response time can result in communication delays between vehicles and infrastructure, which in turn can have adverse effects on the overall performance of the system. These research findings highlight the critical need for enhancing the response time of the ITS-G5 protocol to ensure optimal communication efficiency, particularly in applications related to autonomous driving and road safety.

According to Mayssa Dardouret al. [10] an arrangement has been proposed to accomplish swift response durations and concentrates on the distribution of Cooperative Awareness Messages (CAM) and Decentralized Environmental Notification Messages (DENM) for the enhancement of road user safety. With the intention of averting mishaps, an algorithm for CAM and DENM distribution has been formulated, ensuring prompt notifications in the event of abrupt vehicle obstruction emergencies. Furthermore, a comprehensive and optimized railway braking plan is introduced to further diminish the chances of accidents. This strategy aims to supply effective and timely deceleration of trains, granting road users ample time to clear the level

crossing well in advance and alleviating the potential for potential collisions.

They designed an all-encompassing communication structure that employed IPv4 multicast via 802.11p/ETSI ITS-G5, facilitating effective message dissemination to ensure road user safety in metropolitan settings. Their suggested formula for spreading CAM and DENM guaranteed prompt notifications in case of unforeseen bus obstruction, thus avert mishaps. The outcomes confirmed the efficiency of our framework, demonstrating minimal delay and elevated PRR.

Thomas Otto et al. [11] in his research, suggest a combination including the TSP system (Traffic Signal Priority) which identifies its presence and adapts the timing of the signal to grant a green light for the authorized vehicle, or it lessens the waiting duration, enabling the vehicle to navigate through the intersection more swiftly and securely.

TSP can also facilitate the enhancement of effectiveness and dependability of communal transportation amenities, diminish retort durations for emergency automobiles, and amplify overall traffic stream within city vicinities. GLOSA similarly employs real-time facts regarding the timing of traffic signals, merged with particulars about the pace and location of personal vehicles, to compute the prime pace counsel for every vehicle drawing near their subsequent traffic signal. Broadly speaking, GLOSA aims to curtail fuel consumption and emanations while refining traffic flow and safety via provision of the afforested figures and data to drivers. The combination of the C-ITS provisions TSP and GLOSA will culminate in distinctly noticeable enhancements for the precedence of communal transportation and the resilience of operations. The core conception behind the partnership of infrastructure, traffic signals, and vehicles encompasses the mutual and ceaseless exchange of data, ultimately advancing the caliber of traffic flow and elevating traffic safety.

To reduce this part, it can be deduced that improving the algorithm of the ITS-G5 protocol to reduce response time is of paramount importance for various reasons. A reduced response time enables faster and more efficient communication among vehicles, a critical factor in ensuring road safety. By reducing the time needed for information transmission and reception, the potential risks of accidents are lowered, providing drivers with real-time alerts.

III. THE ITS-G5 COMMUNICATION SYSTEM

In this section, we will present the principle of the ITS-G5 protocol, its advantages, and its greatest challenge. Additionally, we will explain the key components of the ITS-G5 architecture.

A. ITS-G5 (Intelligent Transport Systems)

ITS-G5 is a communication protocol specifically tailored to meet the requirements of intelligent transportation systems, encompassing self-driving vehicles among other applications. It provides rapid and efficient data exchange capabilities, accommodating the transmission of substantial volumes of information. Nevertheless, it's worth noting that the deployment of ITS-G5 is subject to certain limitations, and in

comparison, to alternative protocols, it may exhibit increased latency periods.

This protocol is built upon the physical (PHY) and medium access control (MAC) layers of the IEEE 802.11p standard, which is now incorporated into IEEE 802.11-2016. The PHY and MAC protocols defined by IEEE 802.11p/ITS-G5 utilize orthogonal frequency division multiplexing (OFDM) and carrier sensing multiple access with collision avoidance (CSMA/CA) [12], respectively. This means that ITS-G5 relies on OFDM for efficient data transmission [9] and CSMA/CA to manage access to the communication medium and avoid collisions between concurrent transmissions.

Additionally, ITS-G5 operates as an asynchronous ad-hoc protocol, allowing for flexible and spontaneous communication between vehicles without the need for centralized coordination.

B. Bandwidth Selection

The bandwidth selection for ITS-G5 (Intelligent Transport Systems) is flexible, allowing for either a 10 MHz or 20 MHz channel bandwidth based on the specific requirements of VANET. The ITS-G5 standard incorporates the Geo-Networking protocol to facilitate efficient vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications.

These protocols have been standardized by the ETSI. ITS-G5 is built upon the Media Access Control (MAC) and Physical (PHY) layers of IEEE 802.11p, which are integral parts of the IEEE-802.11-2016 standard. Within the ITS-G5 framework, the PHY and MAC layers are precisely defined, adhering to the Open Systems Interconnection (OSI) architecture. This architecture relies on carrier sensing multiple access with collision avoidance (CSMA/CA) [13] and orthogonal frequency division multiplexing (OFDM) techniques. ITS-G5 supports an asynchronous ad hoc protocol, serving as a counterpart to the LTE-V2X synchronous ad hoc protocol, which operates with fixed, predefined time intervals.

C. ITS-G5 (Intelligent Transport Systems) Challenge

Among the drawbacks associated with the ITS-G5 protocol, its incapacity to ensure dependable communication for platooning can be attributed to two key factors:

1) *Substantial update delay (UD)*: The ITS-G5 protocol showcases a significant lag in updating information shared among vehicles within a platoon. This delay can introduce inaccuracies in the data exchanged [14] among vehicles, consequently compromising coordinated actions and potentially unsafe circumstances.

2) *Limited packet delivery rate (PDR)*: The rate at which data packets are effectively delivered through the utilization of ITS-G5 is notably modest. This deficiency in timely and consistent delivery of critical data can impede the real-time communication essential for efficient platooning.

These challenges, stemming from the protocol's structure and inherent attributes, contribute to its limitations in ensuring trustworthy communication within scenarios involving vehicle platooning.

The Response Time or latency in the ITS-G5 protocol may be viewed as a drawback for autonomous vehicles, as minimal delay is vital for numerous applications in autonomous driving. Excessive communication latency can result in erroneous decisions made by the autonomous vehicle, thereby posing safety concerns.

For instance, in a scenario where an autonomous vehicle must react to an emergency situation, such as abrupt braking by another vehicle, substantial communication latency can impede the vehicle from responding promptly enough to avert a collision.

Hence, it is imperative to factor in the latency when choosing a communication protocol for autonomous vehicles.

D. ITS-G5 (Intelligent Transport Systems) Advantages

Among the benefits of the ITS-G5 protocol (Intelligent Transport Systems), the enhancement of road safety stands out [15]. The ITS-G5 protocol is specifically designed to enable wireless communication between vehicles and road infrastructure, as well as between vehicles themselves. This can positively impact road safety in several ways:

1) *Real-time safety warnings*: ITS-G5 allows vehicles to share real-time information about road conditions, obstacles [16], accidents, and other relevant events. Drivers can be promptly informed about potential hazards, assisting them in reacting appropriately and avoiding accidents.

2) *Collision prevention*: Through direct communication between vehicles, ITS-G5 can aid in collision avoidance using the CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) technique [17], providing proximity warnings when vehicles are about to dangerously approach one another.

3) *Driving assistance*: The information provided by ITS-G5 can enhance driving assistance systems by offering real-time data about traffic [18], weather conditions, and other factors. This enables drivers to adapt their driving more safely and efficiently.

4) *Traffic management*: Authorities responsible for traffic management can utilize ITS-G5 to gather real-time data about traffic patterns, congestion, and road conditions [19]. This information can assist in adjusting traffic signals and implementing traffic management measures to reduce accident risks.

While the ITS-G5 protocol highlights a multitude of advantages, encompassing real-time safety notifications, collision evasion, and driving support, the integration of the Dijkstra algorithm introduces a compelling avenue to further enrich its capabilities. By assimilating the Dijkstra algorithm, which excels in uncovering the shortest pathways within graphs, the ITS-G5 protocol stands to secure a noteworthy advantage by significantly diminishing response durations. Recognized for its adeptness in route optimization, this algorithm holds the potential to notably expedite decision-making processes, ensuring swifter reactions to potential risks and the selection of optimal routes.

In contrast, the innate strengths of the ITS-G5 protocol, spanning real-time safety alerts, mechanisms for collision

prevention, and provisions for driving assistance, establish a robust groundwork for enhanced road encounters, boasting heightened safety and efficiency. Nonetheless, the inclusion of the Dijkstra algorithm ushers in an innovative dimension that capitalizes on its capability to swiftly compute optimal routes. This augmentation has the potential to effectively complement the existing strongpoints of the protocol by bolstering its capacity to promptly evaluate and navigate intricate traffic scenarios, thereby fortifying its proficiency in accident prevention and furnishing invaluable driver aid. Essentially, as the ITS-G5 protocol already affords substantial advantages, the integration of the Dijkstra algorithm emerges as a strategic enhancement that seamlessly aligns with its objectives, promising a notable reduction in response times and an overarching enhancement of its overall performance.

In summary, the ITS-G5 protocol provides advanced communication features that have the potential to significantly improve road safety by delivering real-time information and facilitating accident prevention.

E. ITS-G5(Intelligent Transport Systems) Architecture

The architecture of the ITS-G5 (Intelligent Transport Systems) protocol for self-driving vehicles comprises two primary elements:

- Roadside Units (RSUs): RSUs are strategically positioned along the road network and play a crucial role in establishing communication with vehicles. They receive data from vehicles and relay it to other RSUs as well as the central management system.
- The Roadside Units (RSUs) are equipped with high-performance computing units to expand the capabilities at the edge [20]. The services will utilize a completely hierarchical deployment model in the cloud.
- Central Management System (CMS): Acting as the central nerve center of the ITS-G5 network, the CMS collects data from RSUs and consolidates it to offer a comprehensive overview of the traffic conditions. Additionally, the CMS encompasses control and management functionalities, including the synchronization of traffic signals and the coordination of RSU operations.

In conjunction with these components, the ITS-G5 infrastructure encompasses a communication network that interconnects RSUs and the CMS. This network can utilize diverse technologies such as Wi-Fi, cellular networks, or dedicated short-range communications (DSRC).

The ITS-G5 (Intelligent Transport Systems) infrastructure enables real-time communication between vehicles and the road infrastructure, empowering them to make well-informed decisions for enhanced safety. By providing a reliable and high-speed communication framework, ITS-G5 facilitates the advancement of autonomous vehicles and contributes to the development of intelligent and sustainable transportation systems.

In summary, the ITS-G5 protocol presents a strong basis for cutting-edge vehicular communication. Its advantages encompass heightened road safety, effective traffic control, and

enhanced driver support. Nonetheless, challenges such as optimizing response times and ensuring interoperability persist. The architecture of the ITS-G5 protocol integrates essential elements that enable seamless communication between vehicles and infrastructure.

IV. ITS-G5 (INTELLIGENT TRANSPORT SYSTEM) PROTOCOL ALGORITHM

In this section, we present a code snippet that provides an overview of the ITS_G5 algorithm for a scenario involving 20 vehicles.

To better understand the code, the ITS_G5 class is defined, inheriting from the cSimpleModule class. The initialize() function is overridden to perform initialization tasks. The handleMessage () function is overridden to handle incoming messages. The Define_Module () macro is used to define the module. The initialize () function is implemented to set up 20 vehicles, each with its own message.

A loop is used to create 20 instances of cMessage named "vehicle_msg". The created message is sent through the "out" gate. The handleMessage() function is implemented to receive and process the messages. All vehicles receive and process the message. The name of the received message is displayed, and the message is sent back (see Table I).

Algorithm 1: ITS-G5 Algorithm for Up to 20 Vehicles

```
class ITS_G5 : public
cSimpleModule { protected:virtual void initialize()
override;
virtual void handleMessage(cMessage *msg) override;
};
Define_Module(ITS_G5);
void ITS_G5::initialize() {
// Establish a fleet of 20 vehicles, each equipped with its
individual communication.
For (int i=0; i<20; i++) {
cMessage *msg = new
cMessage("vehicle_msg");
send(msg, "out"); }
void ITS_G5::handleMessage(cMessage *msg){
// All automobiles will receive the communication and handle
EV << "Received message: " << msg->getName() <<
endl;
// All automobiles will receive the
communication and handle it. (msg,
"out");
}
```

TABLE I. SUMMARY OF KEY PARAMETERS AND ACTIONS IN THE PROVIDED ITS-G5 ALGORITHM

Parameter	Description
Class Name	Starting node for Dijkstra's algorithm
Class Inheritance	Inherits from cSimpleModule
Method initialize()	Initializes the module, sets up 20 vehicles with individual messages
Method handleMessage()	Handles received messages, processes and resends them
ModuleDefinition	Define_Module(ITS_G5);
MethodInitialize()	Initializes the module, sets up 20 vehicles with individual messages
Loop inInitialize()	Loops through 20 vehicles, creating and sending messages for each
Send Message	Creates a new cMessage object named "vehicle_msg" and sends it out using the "out" gate
Message Processing	Receives messages, processes them by printing the message name, and resends them using the "out" gate

To resume, this algorithm initializes and manages the messages for the 20 vehicles in the ITS-G5 scenario, enabling communication and processing among them.

V. APPROACH FOR IMPROVING THE ITS-G5 ALGORITHM

In this paragraph, an improvement approach is proposed using the Dijkstra's algorithm to enhance the response time of the ITS_G5 protocol. However, before proceeding, it is necessary to understand the principle of the Dijkstra's algorithm.

A. Dijkstra's Algorithm

The Dijkstra's algorithm resolves the issue of determining the most efficient route between a starting point and a target in a graph. Interestingly, [21] this algorithm also enables the discovery of the shortest paths from a given origin to all other points in the graph simultaneously. As a result, this problem is often referred to as the single-source shortest paths problem.

To understand the Dijkstra algorithm, let's begin with a node referred to the initial vertex. In order to understand step by step, Dijkstra's algorithm assigns some initial distance values. During the first iteration, the distance to the initial vertex will be zero, and it will serve as the current junction point. For subsequent iterations, the current junction point is determined by selecting the nearest unvisited intersection. This process becomes straightforward.

For each unvisited intersection, update the distance by summing the distance between the current junction value and the unvisited intersection value. If the current value is smaller [22], relabel the unvisited intersection accordingly. Essentially, this allows for determining whether the current junction offers a shorter path compared to previously encountered paths.

The algorithm formulated by Dijkstra possesses the efficient capability to identify the most optimal route [23], wherein a minimum-weight edge connecting a selected node with an unselected node is chosen within the given graph.

B. Analysis of the Dijkstra Algorithm for Finding the Shortest Path between Neighboring Points

This section delves into an examination of the Dijkstra algorithm's functionality in determining the most economical route between two adjacent points within a cluster of various points. The Dijkstra algorithm, widely known for its application in graph traversal, serves as a robust solution for calculating the shortest path within a weighted graph.

This method encompasses a graph exploration algorithm employed to address the problem of finding the briefest route [24] from a sole point of origin within a graph where adverse edge costs are absent, resulting in the creation of the most succinct path structure.

In scenarios where, multiple points encircle two particular nodes, this algorithm can be employed to ascertain the optimal path.

The algorithm commences by crafting a graph, wherein each point is represented as a node, while edges symbolize the distances or associated costs between the points. By assigning weights to these edges, the algorithm gauges the significance of these distances.

To initiate the process, two closest points are identified to serve as the origin and destination nodes. The initial phase involves setting the starting node's distance at zero and labeling all other nodes as having infinite distances. These nodes are categorized within an untouched set.

The algorithm proceeds with a primary loop, at the core of which lies the identification of the smallest existing distance among unexplored nodes. At the outset, this could potentially be the starting node. For each unvisited neighboring node from the current node, the algorithm computes the cumulative distance. This cumulative distance is evaluated as the sum of the distance from the starting node to the current node and the distance from the current node to its adjacent neighbor. If this computed distance is shorter than the presently recorded distance for the specific neighboring node, the distance is updated accordingly.

Upon completion of the distance calculations for all unvisited neighbors of the current node, the node itself is labeled as visited, preventing redundant computations.

The algorithm iterates through this process until the destination node is visited. Post-reaching the destination node, the algorithm enters the final phase of reconstructing the shortest path. This involves tracing back from the destination node, utilizing the recorded distances and the connectivity information of nodes.

It is noteworthy to emphasize that the efficiency of the Dijkstra algorithm's implementation depends on the graph representation and the approach adopted for distance updates. Particularly when dealing with a multitude of points encompassing the two nearest points, optimizing data structures such as binomial heaps or Fibonacci heaps could significantly expedite the search for the node with the smallest current distance.

C. Approach of Improvement

The enhancement process could be as follows:

- Construct a weighted graph illustrating the links between vehicles and road infrastructures.
- Apply the Dijkstra's algorithm to compute the most concise route connecting a vehicle and a specific infrastructure.
- Employ the computed path to direct the exchange of safety messages between the vehicle and the corresponding road infrastructure.
- Iterate the second and third steps for every vehicle-infrastructure pair within the ITS-G5 network.

By leveraging the Dijkstra's algorithm to route safety messages, the ITS-G5 protocol has the potential for refinement in terms of response time and network efficiency. Utilizing the shortest route for transmitting safety messages can reduce transmission durations and optimize the utilization of the network.

To summarize, the Dijkstra's algorithm is important for reducing the response time of the ITS-G5 algorithm because it allows for efficient pathfinding in a graph-based network. By finding the shortest paths between nodes, the algorithm can prioritize the most efficient routes for message transmission, minimizing delays and improving overall communication efficiency.

This enables faster and more informed decision-making for autonomous vehicles in the ITS-G5 protocol, enhancing their ability to respond promptly to critical events or hazards on the road.

VI. IMPROVED ALGORITHM OF THE ITS-G5 PROTOCOL

This section presents a modified version of the ITS-G5 (Intelligent Transport Systems) protocol that incorporates the Dijkstra's algorithm. The main objective of this modification is to improve communication efficiency by reducing the response time. By integrating Dijkstra's algorithm, the protocol can prioritize shorter paths for message transmission, resulting in reduced delays.

Algorithm 2: Improved Algorithm of the ITS-G5 Protocol

// Framework to store the node data.

```
Struct Node {
    int node;
    int cost;
};
```

// Procedure to locate the most efficient route between two nodes utilizing Dijkstra's Algorithm.

```
Void dijkstra(int source, vector<vector<int>>& graph,
vector<int>& dist) {
    // Establish a prioritized queue for node storage.
    priority_queue<Node, vector<Node>, greater<Node>> pq;
    // Generate an array of explored nodes.
```

```
vector<bool> visited(graph.size(), false);
// Include the origin node in the prioritized queue with a cost of 0.
pq.push({ source, 0 });
dist[source] = 0;
// Incorporate the initial node into the priority queue with a cost
of 0.

while (!pq.empty()) {
    // Retrieve the present node from the
    prioritized
    queue.int u = pq.top().node;
    pq.pop();
    // Iterate over the adjacent nodes of the current node.

    For (int v = 0; v < graph[u].size(); v++) {
        // Verify if the node has not been traversed and if
        the cost is lower than the current value.

        if (!visited[v] && graph[u][v] != -1 &&
            dist[v] > dist[u] + graph[u][v])
            {
                // Modify the expense of the node.
                dist[v] = dist[u] + graph[u][v];
                // Add the node to the priority queue.
                pq.push({ v, dist[v] });
            }
    }
}
```

TABLE II. IDENTIFICATION OF THE EMPLOYED PARAMETERS

Parameter	Description
'Source'	Starting node for Dijkstra's algorithm
'Graph'	Adjacency matrix representing the graph
'Dist'	Vector storing shortest distances from the source
'Pq'	Priority queue storing nodes for exploration
'Node struct'	Structure to store node information
'U'	Current node being processed
'V'	Neighbor Node of the Current node

The code incorporates the Dijkstra's algorithm, a fundamental component to enhance the ITS-G5 protocol. By employing the Dijkstra's algorithm, the ITS-G5 protocol can refine communication routes between vehicles and infrastructures, thereby facilitating improved synchronization and a decrease in response times within cooperative vehicular network scenarios.

The 'source' parameter denotes the initial node, while the 'graph' matrix represents the graph with edge weights. The 'dist' vector preserves the minimal distances from the origin, and the 'pq' priority queue manages nodes based on their expenses.

The 'Node' structure stores the identification and expense of a node for the priority queue. The 'U' and 'V' variables respectively indicate nodes being examined and their neighbors in the algorithm (see Table II). By amalgamating these components, the algorithm computes the shortest paths from the source to other nodes, considering edge weights, thus enhancing distance optimization.

Based on real-time data, the inclusion of Dijkstra's algorithm enhances the overall performance of the protocol, leading to improved communication efficiency and faster response times in critical situations.

VII. RESULT

Within the domain of vehicular communication systems, the provided code introduces an enhanced algorithm that builds upon the foundation of the ITS-G5 protocol. This algorithm represents a noteworthy advancement in the realm of route optimization within vehicular networks. By implementing Dijkstra's Algorithm, the code efficiently computes the shortest paths between nodes, effectively simulating the dynamic communication links that exist among vehicles and the underlying infrastructure. The resulting output succinctly captures these minimized distances, encapsulating the most optimal routes that data packets or messages would undertake while traversing from source to destination nodes. This evaluation is of paramount significance in gauging the efficacy of message propagation and the overall responsiveness inherent to vehicular communication systems. Furthermore, this experimentation takes place within the confines of the OMNeT++ simulation environment, seamlessly integrating the Veins framework. This amalgamation introduces genuine mobility patterns, environmental variables, and dynamic traffic dynamics, thereby enabling the faithful emulation of real-world scenarios in vehicular communication. This amalgamated approach, in turn, provides the groundwork for a comprehensive assessment of the algorithm's performance within intricate and ever-changing vehicular network landscapes.

A. Validation of Improved Algorithm for Autonomous Vehicle Communication

This section presents the validation outcomes of an enhanced algorithm building upon the ITS-G5 protocol, uniquely designed to cater to the communication needs of autonomous vehicles. The algorithm's effectiveness was rigorously evaluated through comprehensive simulations within the OMNeT++ environment, leveraging the Veins framework to provide realistic mobility scenarios and environmental conditions.

B. Simulation Setup

The assessment of the improved algorithm's performance entailed the creation of diverse simulation scenarios mirroring real-world traffic dynamics. Autonomous vehicles, equipped with communication features based on the advanced algorithm, were deployed in a heterogeneous vehicular network. Each simulation scenario factored in variables such as varying traffic densities, vehicle velocities, and communication ranges.

C. Evaluation Criteria

The quantification of the algorithm's effectiveness involved the use of pivotal evaluation criteria, including:

- **Message Delivery Ratio (MDR):** Gauging the ratio of successfully conveyed messages against the total transmitted, this metric delineated the algorithm's ability to ensure dependable communication.
- **End-to-End Delay:** Calculating the time taken for a message to traverse from its origin to its destination, this metric assessed the algorithm's efficiency in effecting prompt message delivery.
- **Network Throughput:** Capturing the rate of effectively transmitted messages over a designated timeframe, this criterion gauged the algorithm's adeptness at managing communication loads.

D. Results and Analysis

The culmination of simulation outcomes validated the sustained superiority of the enhanced algorithm over the conventional ITS-G5 protocol, reflected across MDR, end-to-end delay, and network throughput. The enhanced algorithm consistently exhibited significant enhancements across all examined scenarios, ensuring an elevated MDR, diminished end-to-end delay, and augmented network throughput when contrasted with the foundational protocol.

Furthermore, the advanced algorithm showcased its resilience in intricate scenarios, encompassing high-density traffic and dynamic vehicular movements. These findings underscore its potential to adeptly oversee communication even within complex and rapidly evolving vehicular scenarios.

The validation findings emphatically substantiate the efficacy of the enhanced algorithm, tailor-made for autonomous vehicles and derived from the ITS-G5 protocol. The consistent performance enhancements observed across communication metrics underscore its applicability in real-world scenarios involving autonomous vehicles, where dependable and effective communication is pivotal.

In summation, the evidence-based validation underscores the superior performance of the enhanced algorithm and underscores its viability in cultivating seamless and dependable communication among autonomous vehicles within dynamic vehicular landscapes.

VIII. DISCUSSION

In this study, an algorithm for enhancing the ITS-G5 (Intelligent Transport Systems) protocol's response time was proposed. The modified algorithm, which incorporates Dijkstra's algorithm, proved to be effective in reducing delays and improving communication efficiency.

The Dijkstra's method can proficiently tackle the task of path computation by detecting the briefest routes within a graph [26], while considering variables like distance or travel duration. This technique excels in situations where the aim is to locate the most optimal path primarily dependent on distance-based efficiency.

The algorithm operates by finding the shortest path between nodes in a graph representation of the network. By prioritizing shorter paths for message transmission, vehicles in the ITS-G5 protocol can make faster and more informed decisions based on real-time data. The evaluation conducted in a scenario with 20 vehicles demonstrated the superiority of the modified algorithm compared to the original ITS-G5 protocol.

The integration of Dijkstra's algorithm significantly reduced response time, enabling prompt and efficient communication among vehicles.

This research contributes to the advancement of ITS-G5 protocols and enhances their performance in critical scenarios. Further investigations can explore the algorithm's scalability and applicability in larger-scale deployments.

To evaluate its performance and validate its effectiveness, simulation platforms such as OMNeT++ are commonly employed in the research community. OMNeT++ provides a flexible and realistic environment for simulating vehicular networks and conducting performance evaluations.

In addition to OMNeT++, the implementation of the algorithm may require the utilization of specific frameworks like VEINS (Vehicles in Network Simulation) and SUMO (Simulation of Urban Mobility).

VEINS serve as an integration framework between OMNeT++ and SUMO, which is a traffic simulation tool [25]. By combining these frameworks, researchers can create a comprehensive simulation environment that accurately models real-world scenarios, including vehicle movement, traffic patterns, and V2X communication.

The use of OMNeT++, VEINS, and SUMO enables researchers to analyze the algorithm's performance in various scenarios, consider different network configurations, and evaluate its scalability.

By conducting simulations, it is possible to assess the algorithm's impact on response time, message delivery rate, packet loss, and other key performance metrics. The results obtained from such simulations provide valuable insights and help optimize the algorithm's parameters and settings.

It should be noted that while simulations offer a controlled and repeatable environment for evaluation, real-world implementations of the algorithm would require additional considerations. Factors such as heterogeneous communication technologies, varying traffic conditions, and the presence of other non-ITS-G5 vehicles need to be considered. Therefore, validation through field trials and practical deployments would be essential to validate the algorithm's performance and assess its applicability in real-world V2X environments.

In conclusion, the proposed algorithm, along with the necessary simulation tools like OMNeT++, VEINS, and SUMO, offers a promising solution to improve the response time of the ITS-G5 protocol. Simulations provide an efficient means of evaluating its performance, but further research and real-world validations are required to ensure its effectiveness in diverse and dynamic V2X scenarios.

IX. CONCLUSION

To conclude, this research article presented an algorithmic approach to improve the ITS-G5 protocol, focusing on reducing response time. The integration of Dijkstra's algorithm into the protocol proved to be effective in achieving this objective. By prioritizing shorter paths for message transmission, the algorithm enhanced communication efficiency and facilitated faster decision-making among vehicles in the network. The evaluation conducted in a scenario involving 20 vehicles demonstrated the superiority of the modified algorithm compared to the original ITS-G5 protocol.

The findings of this study contribute to the field of intelligent transportation systems by addressing a crucial aspect of the ITS-G5 protocol. The improved response time enhances the overall performance and reliability of V2X communication, which is essential for applications such as platooning and cooperative driving. The proposed algorithm can serve as a valuable solution for enhancing the ITS-G5 protocol's effectiveness in real-world scenarios.

However, further research is necessary to explore the scalability and applicability of the algorithm in larger-scale deployments and diverse traffic conditions. Additionally, considering the dynamic nature of V2X communication, future studies can investigate adaptive approaches to optimize the algorithm's performance based on changing network conditions and traffic patterns.

Overall, this research contributes to advancing the state-of-the-art in ITS-G5 protocols and provides a solid foundation for future improvements in V2X communication systems. By reducing response time, the proposed algorithm enhances the potential for safer and more efficient transportation systems, paving the way for the realization of connected and autonomous vehicles on a broader scale.

REFERENCES

- [1] K. Abboud, H. A. Omar, and W. Zhuang, "Interworking of dsrc and cellular network technologies for v2x communications: A survey," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9457–9470, Dec. 2016.
- [2] V. Mannoni, V. Berg, S. Sesia and E. Perraud, "A Comparison of the V2X Communication Systems: ITS-G5 and C-V2X", Published in: *IEEE 89th Vehicular Technology Conference (VTC2019- Spring)*, June 2019.
- [3] P. Roux, S. Sesia, V. Mannoni, & E. Perraud, "System Level Analysis for ITS-G5 and LTE-V2X Performance Comparison." *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. doi:10.1109/mass.2019.
- [4] M. Jutila, J. Scholliers, M. Valta, K. Kujanpää, "ITS-G5 performance improvement and evaluation for vulnerable road user safety services", *IET Intelligent Transport Systems, Selected Papers from the 22nd ITS World Congress*, March 2017.
- [5] J. van Dam, N. Bißmeyer, C. Zimmermann & K. "tEckert, Security in Hybrid Vehicular Communication Based on ITS G5, LTE-V, and Mobile Edge Computing", *Fahrerassistenzsysteme*, 80–91. doi:10.1007/978-3-658-23751-6_8, 2018.
- [6] B. Fernandes, J. Rufino, M. Alam & J. Ferreira, "Implementation and Analysis of IEEE and ETSI Security Standards for Vehicular Communications", *Mobile Networks and Applications*, 23(3), 469–478. doi:10.1007/s11036-018-1019-x, Feb 2018.
- [7] M. Karoui, A. Freitas; G. Chalhoub, "Performance comparison between LTE-V2X and ITS-G5 under realistic urban scenarios"

- [8] I. Rachdan, S. Sand, "ITS-G5 Challenges and 5G Solutions for Vehicular Platooning", IEEE 2016.
- [9] P. Pagana, "Development of an ITS-G5 Station, from the Physical to the MAC Layer", Intelligent Transportation Systems, Pages 37, 2016.
- [10] M. Dardour, M. Mosbah and T. Ahmed, "Improving Emergency Response: An In-Depth Analysis of an ITS-G5 Messaging Strategy for Bus Blockage Emergencies at Level Crossings", Journal of Network and Systems Management, 2023.
- [11] T. Otto, I. Partzsch, J. Holfeld, M. Klöppel-Gersdorf, V. Ivanitzki, "Designing a C-ITS Communication Infrastructure for Traffic Signal Priority of Public Transport", Appl. Sci., 13, 7650. <https://doi.org/10.3390/app13137650>, 2023.
- [12] A. Bazzi, A. Zanella, I. Sarris, V. Martinez, "Co-channel Coexistence: Let ITS-G5 and Sidelink C-V2X Make Peace", IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), 2020.
- [13] M. Naem Tahir, M. Katz, "Heterogeneous (ITS-G5 and 5G) Vehicular Pilot Road Weather Service Platform in a Realistic Operational Environment", <https://doi.org/10.3390/s21051676>, Sensors March 2021.
- [14] I. Rashdan, F. de Ponte Müller, S. Sand, "ITS-G5 Challenges and 5G Solutions for Vehicular Platooning", Proceedings of the WWR37, 2016
- [15] M. Karoui, A. Freitas, G. Chalhoub, "Performance comparison between LTE-V2X and ITS-G5 under realistic urban scenarios", IEEE 91st Vehicular Technology Conference (VTC2020-Spring) - Antwerp, Belgium (2020.5.25-2020.5.28) 2020.
- [16] M. Naem Tahir, K. Maenpaa, T. Sukuvaara, P. Leviakangas, "Deployment and Analysis of Cooperative Intelligent Transport System Pilot Service Alerts in Real Environment", Published in: 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), 2020.
- [17] A. Bazzi, A. Zanella, Ioannis Sarris, V. Martinez, "Co-channel Coexistence: Let ITS-G5 and Sidelink C-V2X Make Peace", Published in: IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), 2020.
- [18] R. Riebl, C. Obermaier, S. Neumeier, C. Facchi, "Vanetza: Boosting Research on Inter-Vehicle Communication", Proceedings of the 5th GI/ITG KuVS Fachgespräch Inter-Vehicle Communication, 2017.
- [19] A. Turley, K. Moerman, A. Filippi, V. Martinez, "C-ITS: Three observations on LTE-V2X and ETSI ITS-G5—A comparison", NXP Semiconductors White, 2018
- [20] J. Marquez-Barja, B. Lannoo, Dries Naudts, B. Braem, C. Donato, V. Maglogiannis, S. Mercelis, R. Berkvens, P. Hellinckx, M. Weyn, I. Moerman, S. Latre, "Smart Highway: ITS-G5 and C-V2X based testbed for vehicular communications in real environments enhanced by edge/cloud technologies", EuCNC2019, the European Conference on Networks and Communications, 2019.
- [21] M. Adeel Javaid, "Understanding Dijkstra's Algorithm", SSRN, Oct 2013.
- [22] N. Makariye, "Towards Shortest Path Computation using Dijkstra Algorithm", Published in: International Conference on IoT and Application (ICIOT), May 2017.
- [23] R. Dedi Gunawan, Riduwan Napianto, R. Indra Borman, I. Hanifah, "IMPLEMENTATION OF DIJKSTRA'S ALGORITHM IN DETERMINING THE SHORTEST PATH (CASE STUDY: SPECIALIST DOCTOR SEARCH IN BANDAR LAMPUNG)", International Journal Information System and Computer Science (IJISCS), Published 24 December 2019.
- [24] Dian Rachmawati, and Lysander Gustin, "Analysis of Dijkstra's Algorithm and A* Algorithm in Shortest Path Problem", Journal of Physics: Conference Series 1566 (2020) 012061 IOP Publishing, ICCAI 2019.
- [25] G. Qing, Z. Zheng, X. Yue, "Path-planning of Automated Guided Vehicle based on Improved Dijkstra Algorithm", Published in: 29th Chinese Control And Decision Conference (CCDC), 2017.
- [26] C. Sommer, D. Eckhoff, A. Brummer, Dominik S. Buse, F. Hagenauer, S. Joerer & M. Segata, "Veins: The Open Source Vehicular Network Simulation Framework", Chapter, 3161 Accesses, 76 Citations, 2019.

Design and Application of an Automatic Scoring System for English Composition Based on Artificial Intelligence Technology

Fengqin Zhang

School of Foreign Languages, Zhengzhou Tourism College, Zhengzhou 450000, China

Abstract—The automatic grading of English compositions involves utilizing natural language processing, statistics, artificial intelligence (AI), and other techniques to evaluate and score compositions. This approach is objective, fair, and resource-efficient. The current widely used evaluation system for English compositions falls short in off-topic assessment, as subjective factors in manual marking lead to inconsistent scoring standards, which affects objectivity and fairness. Hence, researching and implementing an AI-based automatic scoring system for English compositions holds significant importance. This paper examines various composition evaluation factors, such as vocabulary usage, sentence structure, errors, development, word frequency, and examples. These factors are classified, quantified, and analysed using methods such as standardization, cluster analysis, and TF word frequency. Scores are assigned to each feature factor based on fuzzy clustering analysis and the information entropy principle of rough set theory. The system can flexibly identify composition themes in batches and rapidly score English compositions, offering more objective and impartial quality control. The goal of the proposed system is to address existing issues in teacher corrections and evaluations, as well as low self-efficacy in students' writing learning. The test results demonstrate that the system expands the learning material collections, enhances the identification of weak points, optimizes the marking engine performance with the text matching degree, reduces the marking time, and ensures efficient and high-quality assessments. Overall, this system shows great potential for widespread adoption.

Keywords—English composition; automatic scoring; artificial intelligence; text matching degree; natural language processing

I. INTRODUCTION

For a long time, college English writing teaching has been a 'short board'. The traditional teaching mode of college English writing is simple: teachers construct a writing framework-analysis model, students imitate writing, and teachers provide correct-writing comments. The disadvantages of this model are particularly evident in English as a public class. Correcting compositions is energy-consuming; the teachers are powerless in the latter part of writing teaching [1]. As time passes, the writing teaching link becomes 'top-heavy', and the writing evaluation class often becomes a general model essay appreciation class. Students have a low sense of self-efficacy in English writing learning and even fear and anxiety, and their writing ability does not improve. These 'stubborn diseases' are particularly evident among the non-English majors at the author's independent university. The

students' English foundation is weak, and they have writing problems, but many of them do not know how to improve. This 'old and difficult' problem of college English writing needs to be solved with new ideas. In automatic composition grading, statistics, natural language processing, artificial intelligence and other technologies are used to evaluate and grade compositions. The procedure of automatic composition scoring is to preprocess an English article with initial word segmentation, clauses and part-of-speech tagging, then analyse its morphology, grammar, content richness and other characteristics, and finally score the composition according to appropriate standards [2]. Automatic composition grading brings speed and high efficiency and can greatly reduce teachers' efforts for essay scoring, which leaves them more time to teach students. Simultaneous targeted training can also enable English learners to improve their English writing ability and, when applied to large-scale tests, can greatly reduce manpower and material resources, improve efficiency, and guarantee the impartiality scores. Technology in the field of education, particularly artificial intelligence, comprehensively and profoundly influences the education concept, teaching model and exam method. Based on speech technology in English listening, oral tests have a wide range of applications. Handwriting recognition and natural language understanding, such as artificial intelligence technology, are also being explored and applied in education examination evaluations. Therefore, it is of practical significance to explore an automatic scoring system that is suitable for nonnative English speakers and has low cost, automation and high accuracy [3-5].

Automatic composition grading involves the use of statistics, natural language processing, artificial intelligence and other technologies to evaluate and grade essays. Automatic composition grading has been widely applied to various examinations; indeed, teaching in the classroom has obtained a certain effect, but in the process of using an automatic scoring system, it has been found that even if the input of an article has nothing to do with the thesis topic, as long as there are not too many grammar and vocabulary errors and high scores can be obtained, researchers can realize the importance of track detection to the thesis [6]. Since the 1960s, with the rapid development of natural language processing technology, automatic composition scoring systems have made great progress and have gradually been introduced for a variety of teaching and exam uses. The PEG scoring system does not use natural language processing technology, there is

no study of the composition or content of the chapter structure, and the theme of the thesis is not considered [7]. The IEA scoring system effectively grades sample essays on a variety of topics, automatically judges the content and quality of essays and provides quick feedback. The E-rater grading system, with its complex feature engineering, can better reflect the quality of composition, so its scoring results are highly consistent with those of manual grading. IntelliMetric uses standardized scoring rules and follows the human brain's judgment of points to extract relevant characteristics of essays and then grades essays based on a constructed model [8]. The BETSY scoring system can first extract the characteristics of the quality of a composition and then, according to these characteristics, the thesis is divided into several different levels of people. After manual annotation data are used as the training sample set, the classification of the training to obtain the composition model needs to follow the same method to test the composition, and the extracted features of a classification model can be assigned to the corresponding collection. The above systems mainly use regression and classification methods to score compositions. In recent years, with the rapid development of deep learning technology, many scholars have attempted to use neural networks to grade English compositions [9, 10]. At present, the composition of the track detection method has a certain effect but also has obvious problems: based on the supervised method, the accuracy is higher, but in daily teaching use, it is flexible in composition. Once the system does not include the new theme in the corpus proposition, track detection accuracy will be discounted, and this method is only applicable to large tests. Based on the unsupervised method, user operation is relatively simple and can also adapt to a variety of different propositions, but the current accuracy is not high and cannot meet the requirements of use [11]. Feature extraction may affect the selection of the training model in the process of constructing an automatic composition scoring model. Therefore, the interaction between these two aspects should be fully considered in the process model construction to select a more appropriate feature extraction method and training model. Most automatic composition grading models can achieve good results, but in supervised learning, especially when the number of candidates is large, much effort is needed for composition annotations, and at the same time, the training process is often dependent on the composition title information and has widespread migration problems, so automatic composition grading also needs to be further examined [12-15].

At present, although the momentum of the development and application of artificial intelligence auxiliary to English writing is good, there are few studies in this field on how to combine students' autonomous learning and improve the enthusiasm of students practising writing outside the classroom for autonomous learning, lifelong learning, and learning views; therefore, it is necessary to conduct further research on the English composition score method [16]. This paper discusses how to apply the machine automatic scoring scheme to composition scoring more effectively to ensure the objectivity, fairness and accuracy of composition scoring in various English exams. This paper examines an automatic scoring system based on artificial intelligence technology English composition design and application with the goal of

exploring the teaching effect of improving English writing classes, improving students' self-efficacy and innovation and cultivating their autonomous learning and consciousness of lifelong learning. Ensuring that students' English composition examination papers are high quality is a good objective and worth reviewing, which will provide a rigorous basis for teaching improvement and talent selection.

Despite the evolving landscape of English teaching and grading, there remains a critical gap, that of the effective and efficient application of modern day technologies, such as artificial intelligence, in the field of English composition writing. English composition education, especially among non-English majors at independent universities, has long suffered from the problems of traditional rote learning, lack of individualized attention and ineffective qualitative assessment. Moreover, the emphasis on understanding student self-efficacy and motivation for autonomous learning in the context of AI-led teaching has been underexplored. This paper seeks to elucidate this understudied area, with a specific focus on nonnative English speakers.

The above work mainly discusses the application of an automatic English composition scoring system based on artificial intelligence technology in college English writing teaching. There are some defects in the traditional teaching mode; for example, correcting compositions consumes teachers' energy and leads to an imbalance in the writing teaching process. Therefore, the introduction of an automatic grading system can improve the efficiency and quality of writing teaching. In general, this paper first provides a brief overview of the application of artificial intelligence-based automatic scoring systems in college English writing teaching. At the same time, it also highlights some problems that need further research and improvement and emphasizes the importance of improving teaching effects and cultivating students' self-learning consciousness through this technical means.

II. TECHNICAL THEORY OF AUTOMATIC COMPOSITION GRADING

Automatic composition scoring mainly involves using natural language processing technology to process composition text content and employing statistical methods for analysis and prediction. Among the many types of questions on an English test, English composition questions can most comprehensively reflect an examinee's comprehensive ability to use English. Through writing, examinees can express and transmit their thoughts and opinions and teachers can examine students' logical thinking and language expression ability.

A. An Overview of Related Natural Language Processing Techniques

Word segmentation and clause segmentation are the basis of natural language processing. English word segmentation is relatively simple, generally using spaces and punctuation for natural separation but also a small number of abbreviations. Clause segmentation refers to an article being divided into a single sentence, usually according to punctuation. The technology of English word segmentation and clause

segmentation is currently quite perfect, and there are many open technologies that are very effective in use. A part of speech is the basic grammatical attribute of a word and is generally called part of speech. Part of speech tagging refers to considering the grammatical category of the vocabulary in a sentence, noting its parts of speech and marking it. The common parts of speech of words are nouns, adjectives, adverbs, verbs and so on. Part-of-speech tagging is one of the basic problems of NLP because data preprocessing of many tasks in NLP requires part-of-speech tagging. Word form reduction is an important part of text preprocessing. In English, it generally refers to the restoration of words in any form to the general tense [17,18]. To put it simply, word restoration removes the affixes of a word and keeps only the main part of the word. Generally, restored words also exist in the dictionary. Form reduction is similar to stem extraction, but it is possible that the word from stem extraction does not appear in the dictionary [19].

Generally, English composition questions have the following characteristics:

1) *Lexical features*: English has strict requirements on the form of words; for example, there are clear usage scenarios for singular and plural nouns.

2) *Phrasal features*: English phrases have many types, such as noun phrases, verb phrases, and prepositional phrases. The collocations of these phrases generally have fixed forms and become modules that constitute sentences. This modular structure ensures the dominant characteristics of English forms.

3) *Syntactic features*: The basic sentence structure is subject and predicate, but English also has complex clauses to express rich content, and the existence form of clauses is flexible.

4) *Structural features*: Cohesion between sentences and paragraphs is the basis of coherence in English composition, and coherence is an important prerequisite for rigorous structure.

The general block diagram of the model is shown in Fig. 1. Given the characteristics of a composition, we should consider the following:

1) *Vocabulary of the composition*: According to the number of novel words in the composition, the number of backbone words is used to evaluate whether students' vocabulary is qualified, and according to the spelling of words, their memory of the vocabulary is evaluated.

2) *Composition syntax*: The use of complex clauses is an important indicator of an examinee's command of English.

3) *Whether the composition structure is rigorous*: The main test for students before and after the description needs to be logical, and the context needs to be appropriate.

4) *Composition content*: This mainly investigates whether the composition content is rich and closely related to the topic.

Therefore, we can find English essay scoring with a strong subjectivity, while a large current test will be equipped with corresponding criteria for English composition, but the scoring

criteria from the linguistics angle only provide guidance and are not combined with specific question operability, which is given a set of scoring rules specific to the process of evaluation [20,21]. Careful evaluation and discussion by highly specialized experts is often needed to determine operational scoring criteria.

B. Evaluation Methodology

The Pearson correlation coefficient is a common linear correlation coefficient that can be used to measure the degree of linear correlation between two variables. For linear variables M and N , for example, the Pearson correlation coefficient can be used to measure the related degree. For x , the value of x ranges from $-1 \sim 1$, and the absolute value of x tends to be closer to 1, showing that the strength of the correlation of M and N is closer to zero, indicating a weaker correlation. A positive x suggests a positive association between M and N , and a negative x indicates a negative correlation [22]. Pearson's correlation coefficient is equal to the covariance between linear variables M and N divided by the product of their standard deviations, as shown below:

$$\rho_{xy} = \frac{\text{cov}(M, N)}{\sigma_M \sigma_N} = \frac{E(M - \mu_M)(N - \mu_N)}{\sigma_M \sigma_N} \quad (1)$$

First, calculate the standard deviation and covariance of samples, that is, calculate the Pearson correlation coefficient between samples, denoted as r :

$$r = \frac{\sum_{i=1}^n (M_i - \bar{M})(N_i - \bar{N})}{\sqrt{\sum_{i=1}^n (M_i - \bar{M})^2} \sqrt{\sum_{i=1}^n (N_i - \bar{N})^2}} \quad (2)$$

r can also be estimated by means of the average standard score of sample points, and the calculation result is equivalent to Eq.(2):

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{M_i - \bar{M}}{\sigma_M} \right) \left(\frac{N_i - \bar{N}}{\sigma_N} \right) \quad (3)$$

where, $\frac{M_i - \bar{M}}{\sigma_M}$, \bar{M} and σ_M are the standard score, sample mean, and sample standard deviation for sample M_i , respectively.

In the process of evaluating the automatic scoring system of composition, we mainly compare the scoring of the system with the manual scoring. In this model, the following three indicators are mainly referenced: the average error of scoring, the average accuracy of scoring, and the relevance of scoring [23]. The formulas are as follows:

$$\text{AverageScoreError} = \frac{\sum_{i=1}^N |x_i - y_i|}{N} \quad (4)$$

$$\text{AverageScoreAccuracy} = \left(\frac{\sum_{i=1}^N |x_i - y_i|}{y_i} \right) \frac{1}{N} \quad (5)$$

$$\text{ScoreRelevance} = \frac{N \sum_{i=1}^N x_i y_i - \sum_{i=1}^N x_i \sum_{i=1}^N y_i}{\sqrt{N \sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i \right)^2} \sqrt{N \sum_{i=1}^N y_i^2 - \left(\sum_{i=1}^N y_i \right)^2}} \quad (6)$$

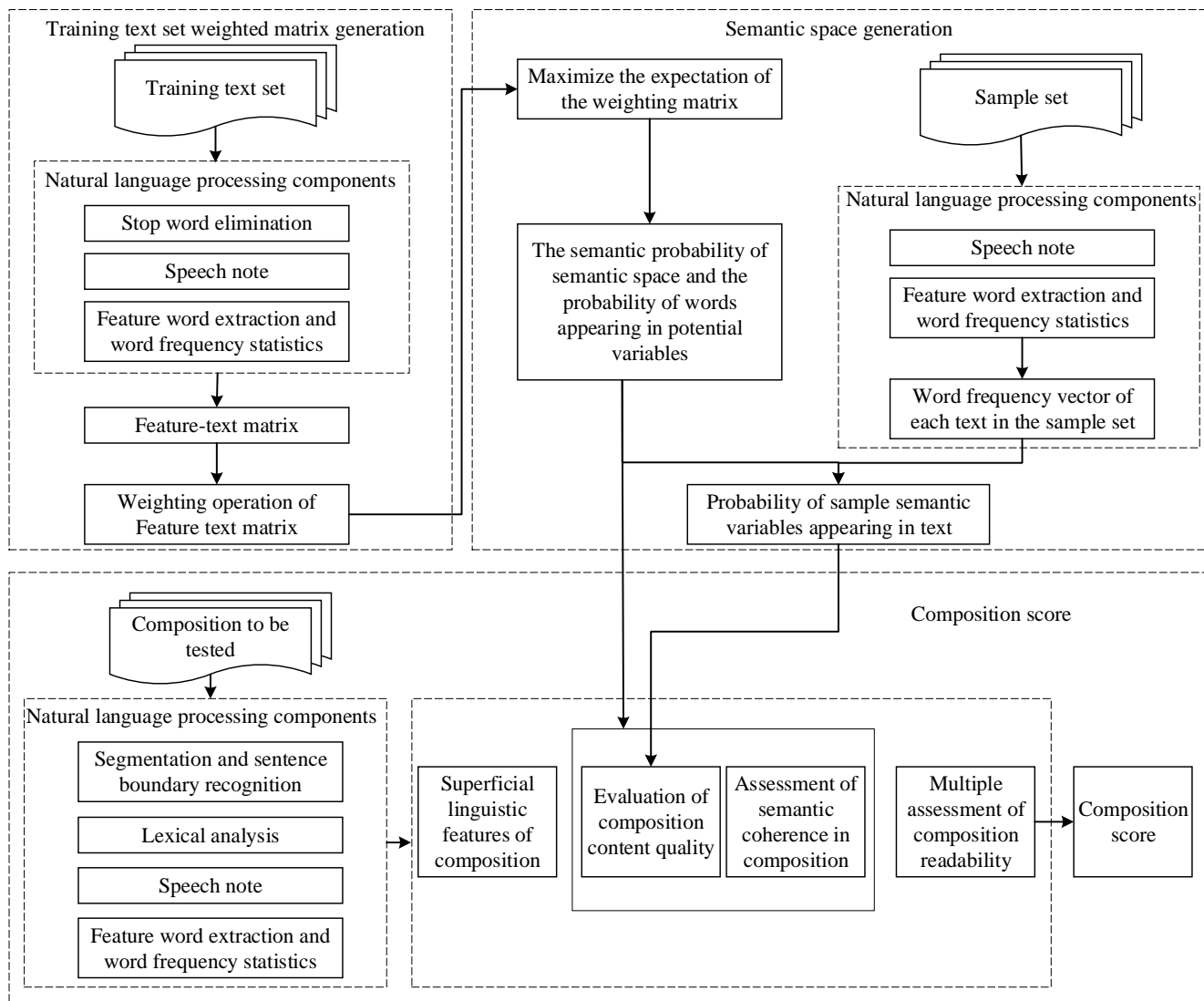


Fig. 1. General block diagram of the model.

C. Neural Network Applications in Natural Language Processing and Composition Grading

Neural networks have always been a research hotspot in artificial intelligence, cognitive science, nonlinear dynamics and other related fields. Neural networks have been used for academic research in recent years. Over the course of these studies, hundreds of neural network models, including pattern recognition, associative memory, signal processing, control engineering, expert systems, combinatorial optimization, image processing and computer graphics, and many other aspects, have been proposed and have made remarkable progress [24]. At present, deep learning has also been introduced in the natural language field, using the concept of word vectors to enable neural networks to complete the work accomplished in the field of statistics. It can be said that neural networks have begun to show advantages in the field of natural language processing. In the process of automatic composition scoring, the biggest problem that puzzles researchers is how to construct a reasonable model for the characteristics obtained from statistical quantification so that it

can evaluate the composition quality well. Because there are many factors that determine the quality of a composition, many different text features can be extracted when the text is processed. These features are complex and changeable, and it is difficult to model them with traditional mathematical theory. Through the understanding of the basic principle of neural networks, combined with the relevant research on automatic scoring technology of composition, the features extracted from composition are processed by artificial neural networks, and the learning, associative memory and distributed parallel information processing functions of neural networks are used to simulate the thinking mode of the human brain [25,26]. Therefore, the constructed neural network model can acquire, learn and reason the expert-rated experience from a large number of expert-rated articles, determine the relationship between the extracted features and the composition score, which is called learning the expert-rated experience, and then grade the text through the learned experience.

After clustering, the word vector has a certain representativeness. Suppose that in the English composition

analysis, after a clustering algorithm has been used to obtain a text clustering, each word in this sentence is in the text clustering, and each different semantic information in the text collection can obtain a word weight value such as frequency and frequency of use [27]. In this paper, by using the automatic grading system of the main statistical, the three characteristics of the term vectors, including word frequency location, size and distribution, the distribution is used to express the sentence, the complexity and diversity of characteristics; for example, in a semantic statement, the author uses more than one word for a description because the authors have a better command of the English language. Based on the above research ideas, this paper extracts and classifies the text features of English compositions.

In the process of grading, teachers can score by item. Usually, there are several indices in grading that are extracted from this model and include the following: score of content quality, semantic coherence, superficial linguistic features of text, and text readability indices. Among them, the semantic score has the largest proportion and indicates whether the central idea of the composition is distinctive and meets the requirements of the topic. If a composition is off topic, then it will have few points. Semantic coherence is also very important; it indicates that the composition of the thought content is consistent, there is a natural smooth transition between statements, text is characterized by shallow linguistics statistics in the text, the word count, the number of sentences, paragraphs and the number of complex words are appropriate, and the main purpose is to evaluate the level of the students to master words, writing skills, etc. The readability index of a text indicates that the text is worth reading. This model uses the features of the existing text to score the text comprehensively.

III. DESIGN AND IMPLEMENTATION OF AN AUTOMATIC SCORING SYSTEM FOR ENGLISH COMPOSITION

A. Requirement Analysis of System

Proposition composition is very common in the daily teaching of a thesis topic. This article designs rating systems, mainly for the proposition composition rate, so the rating system needs to consider not only the composition, such as vocabulary, grammar, and sentences, but also whether the content of the thesis tracks. At the same time, the system's performance and ease of use are also very important [28]. The functional requirements of the system are as follows:

1) The system can easily input composition content and can be repeatedly modified and scored.

2) The system can score multiple compositions at one time, which is convenient for teachers to score the whole class compositions.

3) The interface of the system should be as simple and intuitive as possible so that users can use it quickly and conveniently.

4) The system can efficiently identify the topic of the essay and judge whether the content of the essay to be graded fits the topic.

5) The system can flexibly configure the topic of the thesis composition, which is convenient for users when scoring different thesis compositions.

The automatic scoring system based on the above requirements will be divided into seven modules: login information maintenance, sample volume, candidate set generation, experts, screening and grading, model training and essay scoring, password changes and user information management, including the composition of model training and essay scoring automatic grading and result output two functions.

In summary, an expert-assisted automatic scoring process of English composition can be obtained. Its main workflow includes test paper cutting, image processing, sample paper candidate set generation, sample paper screening and scoring, scoring model training, automatic scoring of composition and result output. The automatic scoring process of English composition is shown in Fig. 2.

1) *Test paper cutting*: The technical staff adopts cutting software to batch cut all candidates' answer sheet images in accordance with the established cutting scheme, forming a separate English composition answer image test paper library.

2) *Image processing*: The English composition answer image examination paper library is processed and the composition content in the picture is obtained and saved as a text form, which is output into a file according to a specific format.

3) *Sample paper candidate set generation*: The machine adopts certain technical means to traverse the whole composition database and screen out the sample paper candidate set that can fully reflect the whole examination paper library of each grade of examination paper level.

4) *Sample volume screening and scoring*: Experts screen relatively uniform sample volume sets from sample volume candidate sets and score them.

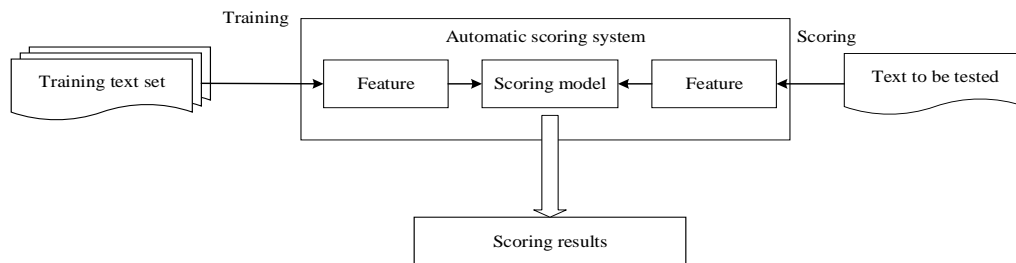


Fig. 2. Automatic scoring process of English compositions.

5) *Training of the scoring model*: The automatic writing scoring model is trained by using the set of selected sample papers and the scores of experts on the set of sample papers as training sets.

6) *Automatic composition scoring*: The automatic composition scoring model obtained by training is applied to paperless marking instead of marking teachers for composition scoring.

7) *Output of scores*: The scores are output.

B. Preprocessing and Feature Extraction for Automatic Grading of English Composition

Before feature extraction of the document, it is necessary to preprocess the composition to some extent to extract feature items. First, it is necessary to divide the composition into paragraphs, sentences, punctuation and words of the text. The obtained paragraphs, sentences, punctuation marks and words are stored separately for the statistical analysis of lexical features and structural features. Natural language processing (NLP) technology is used to tag the parts of speech in the articles to facilitate the statistical analysis of syntactic features and the detection of word errors and grammatical errors. In this paper, the natural language processing tool developed by Stanford University is used for text segmentation and part-of-speech tagging [29].

Composition feature extraction is the core of automatic grading to solve the English composition problem. Because, unlike humans, a computer cannot understand the connotation of a composition or gift and cannot appreciate and evaluate the merits of a composition, it needs to have quantitative data for calculations. Therefore, in the process of implementing automatic grading composition, we need to extract some quantitative data from the text for the computer to calculate and process these data. These data can reflect the real writing level of students. Through these data, we can establish a certain mathematical model, which can be an effective evaluation of students' writing.

The model mainly covers 64 features of composition, such as vocabulary use, phrase collocation, sentence, coherence, organizational structure and fluency, and then divides these features into three types: morphology, syntax and structure. Training sets are used to train the designed neural network [30]. Then, it evaluates the linguistic quality and organizational structure of the composition from the aspects of morphology, syntax and structure. In addition to the evaluation of the above three aspects, this paper also analyses and evaluates the semantic content of the text. Finally, the characteristics of the above aspects are comprehensively scored.

$p(d_i)$ represents the probability of document d_i appearing in the dataset, and $p(z_k|d_i)$ represents the probability distribution of the topic of k document d_i ; $p(w_j|z_k)$ represents the probability distribution of words in topic z_k , each topic

follows a polynomial distribution over all terms, and each document follows a polynomial distribution over all topics. Based on the above probability distribution, a document d_i is randomly selected according to the document probability distribution $p(d_i)$ in the dataset, the topic z_k is selected according to the document topic probability distribution $p(z_k|d_i)$, and then the word of the document d_i is selected according to the keyword probability distribution $p(w_j|z_k)$ of the topic z_k . The data that we can observe are (d_i, w_j) , and z_k represents the implicit variables. The joint distribution of (d_i, w_j) is:

$$p(d_i, w_j) = p(d_i)p(w_j|d_i) \quad (7)$$

$$p(w_j|d_i) = \sum_{k=1}^K p(w_j|z_k)p(z_k|d_i) \quad (8)$$

The $p(z_k|d_i)$ and $p(w_j|z_k)$ distributions correspond to two sets of polynomial distributions. To estimate the parameters of these two sets of distributions, the expectation maximization algorithm should be used. The EM algorithm is divided into E steps and M steps, where the E step is used to solve the post probability distribution of the implicit variable z_k when d_i, w_j is known, and the formula is as follows:

$$p(z_k|d_i, w_j) = \frac{p(w_j|z_k)p(z_k|d_i)}{\sum_{k=1}^K p(w_j|z_k)p(z_k|d_i)} \quad (9)$$

The left side of the formula represents the probability of the occurrence of the k th implied topic under the probability of the occurrence of the i th document and the j th word. Step M is used to solve the posterior probability distribution $p(w_j|d_i)$ of topic words and topic documents when $p(z_k|d_i, w_j)$ are known. The formula is as follows:

$$p(w_j|z_k) \propto \sum_{i=1}^N n(d_i, w_j)p(z_k|d_i, w_j) \quad (10)$$

$$p(z_k|d_i) \propto \sum_{j=1}^M n(d_i, w_j)p(z_k|d_i, w_j) \quad (11)$$

It can be found from the above formula that the E and M steps of the expectation maximization algorithm depend on each other, and the three distributions of , , and can be obtained $p(z_k|d_i, w_j)$ $p(w_j|z_k)$ $p(z_k|d_i)$ after a continuous iterative solution. The scoring process of the composition scoring system is shown in Fig. 3.

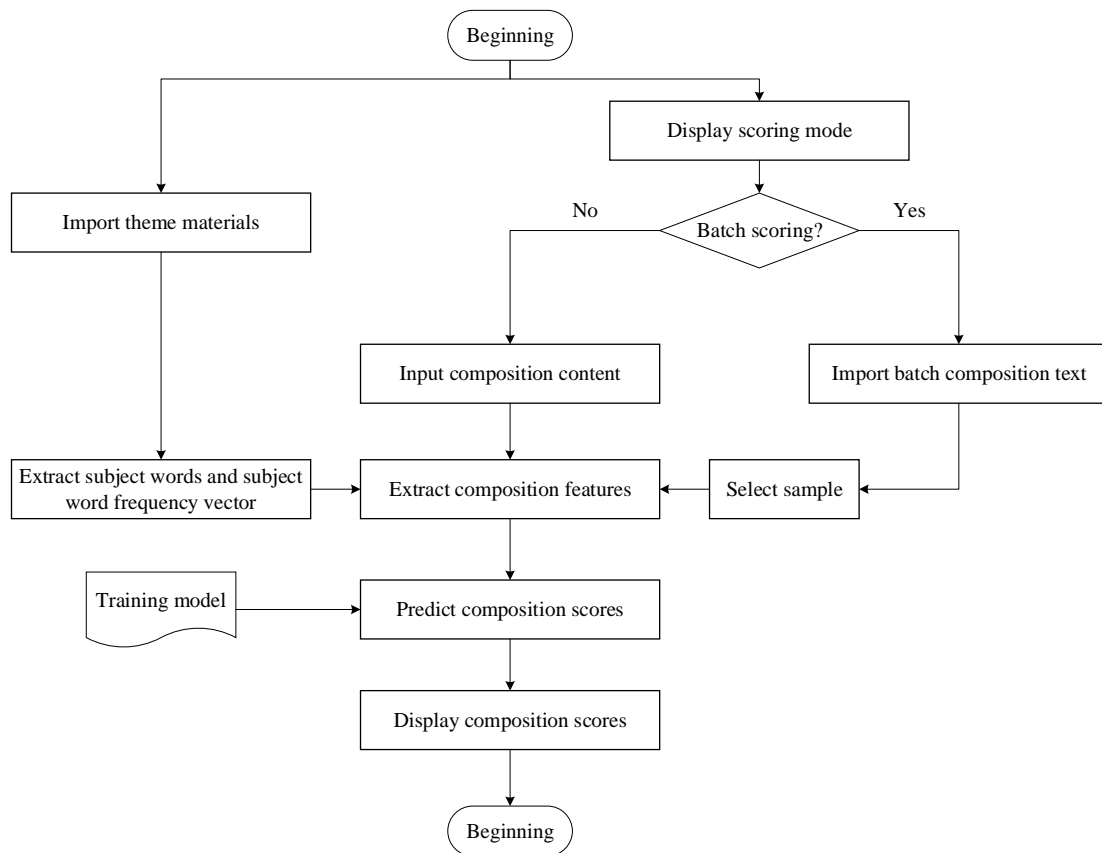


Fig. 3. Scoring process of the composition scoring system.

IV. SYSTEM TEST AND ANALYSIS

A. Evaluation and Verification of an Automatic Scoring System for Composition Based on Word Vector Clustering

To verify the scoring effect of the automatic scoring system based on word vector clustering, this paper takes the standard data provided in a composition scoring competition as the research object, and the total amount of processing of various samples is shown in Fig. 4. In this verification, all the samples except the two parts of the calibration set and the exception answer were scored by a computer. According to the results, the intelligent score of Chinese composition was 420070, accounting for 99.82% of the total sample size, and the intelligent score of English composition was 418820, accounting for 99.53% of the total sample size. The abnormal samples included high similarity with the Chinese text, high similarity with the current test paper (reading comprehension), and high similarity with each other. There were 235 Chinese compositions, accounting for 0.06% of the total test papers. English composition 1469, accounting for 0.35% of the total examination papers. The subject expert group conducts targeted quality inspection re-evaluation on abnormal samples.

In this paper, 235 Chinese compositions and 1469 English compositions were selected and matched with the standard target text. The comparison of the bit error rate of key word recognition is shown in Fig. 5. The statistical results are as follows: the recognition accuracy of Chinese characters is 97.6%, and the recognition accuracy of English words is

97.3%. This high-precision transliteration recognition has three important factors: first, examinees' attention to the composition of the college entrance examination ensures the standardization of writing. Second, the Chinese composition area is designed in square paper format, and the English composition area is designed in a line-by-line underline format to ensure the writing position of characters. Third, there are advanced recognition algorithms. These three factors can ensure the accurate recognition of all the scoring samples, and the overall transliteration recognition rate should be kept at approximately 97%, which can meet the actual requirements of marking papers.

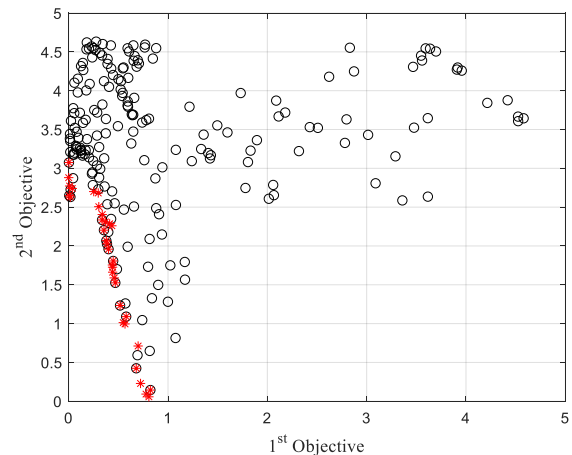


Fig. 4. Total amount of processing of various samples.

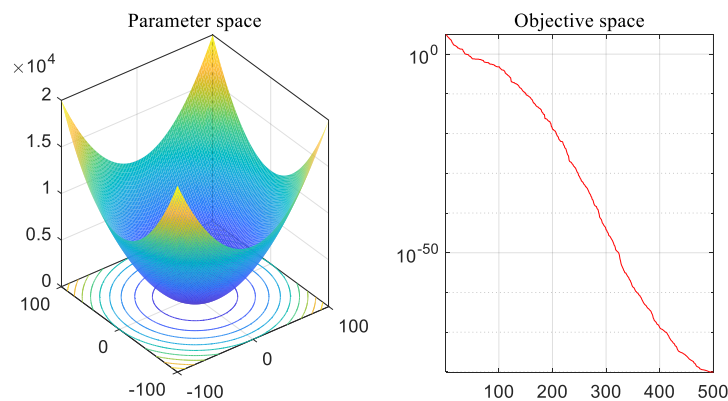


Fig. 5. Comparison of the bit error rate of key word recognition.

B. Evaluation and Comparison of Automatic Scoring Algorithms for Composition

The comparison of the training time of different algorithms is shown in Fig. 6. The average score of the machine score and the average score of the manual score were less than 1 point, and the standard deviation was basically the same. The correlation between the machine score and report score is 0.95, and the consistency rate is 95.24%, which is very close to the correlation and consistency rate of Manual 1 and Manual 2 and is at a high level, which proves that the overall effect of the intelligent score is good. At the same time, it also shows from another angle that the intelligent marking system has a very high learning ability for the calibration set and basically reaches the level of mastering the marking standards of the teachers. It can be seen that most students do not resist the use of correcting network writing, and the correcting network has a good impression. The correcting network writing system makes sense and can stimulate students' desire for writing and continuous improvement. This model provides statistics on the content quality, coherence, readability and basic information of the composition and gives students detailed feedback so that students can better understand their own composition.

As seen from the prediction results of automatic scoring based on the random forest model in the figure, in the scoring results of the composition subset based on the random forest algorithm, the quadratic weighted K value is generally above 0.78, the highest value is 0.905, and the average value is 0.862. The lowest value of weighted K obtained by the international scoring algorithm is 0.654, the highest value is 0.755, and the average value is 0.792. In terms of the prediction results, the calculation method in this paper is obviously better than the existing prediction model, 10%~18% higher than the general algorithm, and can basically achieve the matching effect with the artificial score. Further analysis of the composition sample structure shows that the random forest algorithm based on the bagging method can effectively avoid overfitting error in the case of an insufficient sample size after obtaining accurate clustering vector features, thus reducing the variance value. When the number of samples is less than 1400, the quadratic weighted K value of the conventional model prediction algorithm decreases obviously and is basically lower than 0.7. The average score deviation

rate of each formula combination is shown in Fig. 7.

This model not only gives the composition of the overall score as the composition of the machine but also, in this paper, extracts feature feedback to assist students in better understanding their writing level. This process uses internet technology to achieve a short score. With this composition, students can receive reliable information in a timely and effective manner, which is also one of the advantages of intelligent reading systems. As shown in Fig. 8, the score of the essay given by the machine is 8.41; while the score given by the teacher manually is 9.0, indicating that the consistency between the essay and the manual score is very high. The text coherence is 0.25, 0.77, 0.62 and 1.0, indicating that the text coherence is very good. The readability of the text is 11.89, the comprehensibility is 193.98, and the writing level is 10.82. In the linguistic features of the text, the total number of words is 204, word density is the ratio of the number of different words to the total number of words, word density is 0.66, complex word digit number is 36, sentence number is 17, paragraph number is 3, and average sentence length is 12. All these features can effectively give students an intuitive understanding of their own compositions. Therefore, the students know more about their own composition to improve the efficiency of their learning. The impact of different training sets on model performance is shown in Fig. 8.

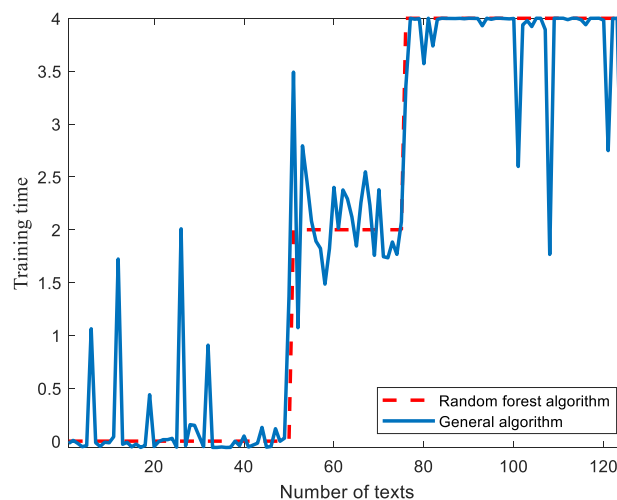


Fig. 6. Comparison of training time of different algorithms.

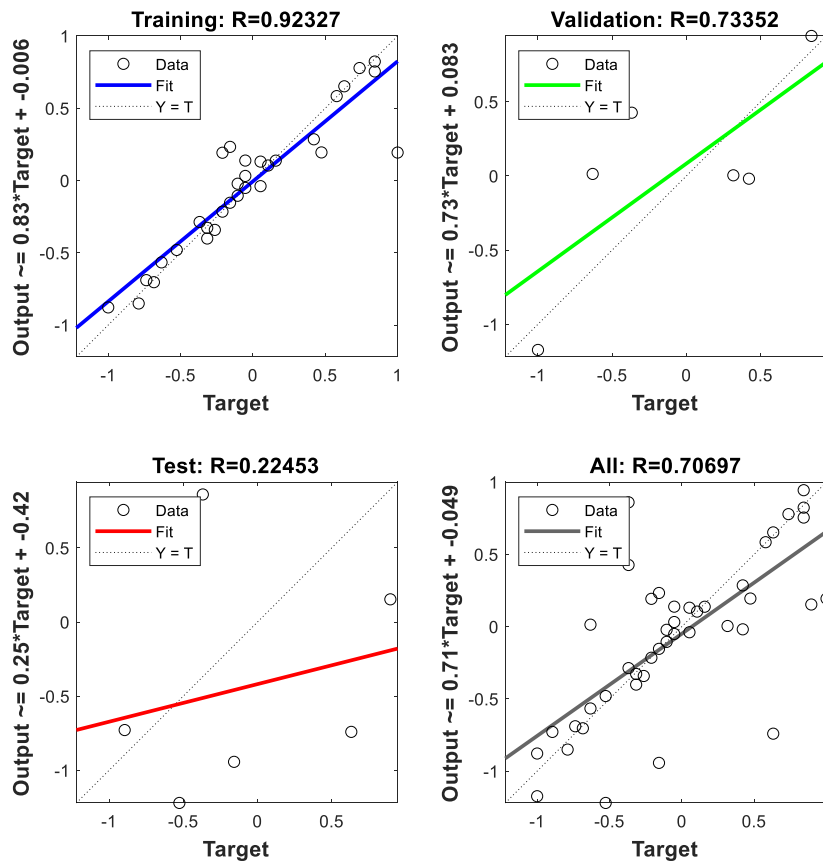


Fig. 7. Average score deviation rate of each formula combination.

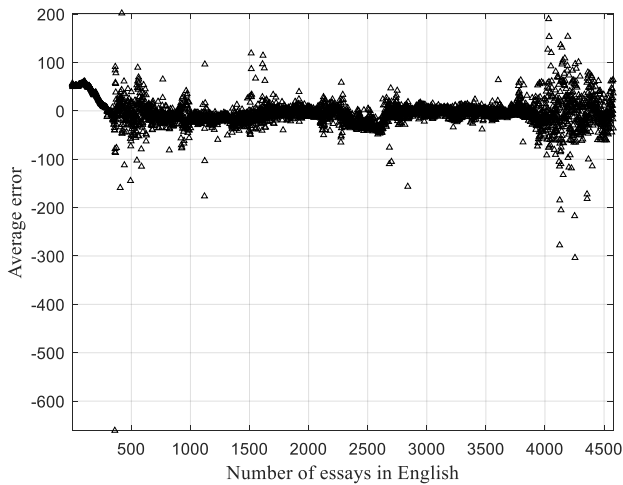


Fig. 8. Impact of different training sets on model performance.

V. ANALYSIS

An AI-based English essay grading system utilizes artificial intelligence technology to automatically evaluate and score essays written in the English language. This advanced system has gained popularity due to its potential to efficiently assess a large number of essays with consistent accuracy, saving time for educators and providing immediate feedback to students. The advantages of an AI-based English essay grading system lie in its consistency, time efficiency, scalability, immediate feedback, objective evaluation,

enhanced learning experience, consistent feedback, and analytics capabilities. While it has limitations in contextual understanding and ethical considerations, integrating this technology with human expertise can lead to a balanced approach that leverages the strengths of both artificial intelligence and human judgment to improve the overall quality of writing assessment and instruction.

VI. CONCLUSION

With the further development of global economic integration and internet technology, English has become the mainstream language for people worldwide to communicate. Therefore, an increasing number of governments are paying attention to the learning of English. Early essay scoring was performed manually, that is, when some teachers read the composition of the paper-based test answers directly to score them, the results are affected by the fact that some teachers' subjective factors are quite large, which is almost impossible to achieve in the large exam score standard consistency requirements, and given how much is required by the evaluation objective, impartiality is more difficult to achieve. Over time, the method of writing assessment has reached the stage of combining intelligent correction with teacher correction. To date, the introduction of artificial intelligence into college English writing teaching can effectively assist teachers in completing composition correction and data analysis more efficiently to a certain extent, encourage students to evaluate, learn from each other more conveniently, and cultivate students' awareness of independent English

writing and lifelong learning. Therefore, this paper examines how to apply an automatic scoring scheme to composition scoring more effectively and provides some reference for research on the automatic scoring direction of English composition under the background of artificial intelligence. The main research contents and conclusions are summarized as follows:

1) Natural language processing technology is used to analyse the text features of a composition, including part of speech tagging, syntactic analysis and article organization structure analysis.

2) Through the correlation analysis of the statistical features, the features related to scoring were extracted, and the relevant features were classified according to the different angles of the features reflecting the quality of the composition.

3) Semantic analysis technology was combined to analyse the semantic content of the composition and obtain the content quality characteristics of the composition.

4) The neural network is successfully applied in the automatic composition scoring process, and the mapping relationship between text features and composition scoring is obtained.

5) A comprehensive score for the composition was given based on its morphology, syntax, organizational structure and content quality.

VII. FUTURE WORK

This paper can not only directly evaluate the language quality and organizational structure of a composition but also evaluate the content quality of the composition and conduct statistical analysis of the lexical use, syntactic use, semantic coherence and readability of the composition. Students and teachers can quickly obtain the quantitative indicators of a composition, providing a powerful reference for teachers and students and not only greatly reducing teachers' workload but also increasing the number of English learners' writing. Due to the large number of English learners in China and the shortage of teacher resources, the English composition scoring system has great application value. Although this model has certain functions, it still has certain deficiencies, which are embodied in the following aspects:

1) In this article, the model used in the training process of neural networks included a Chinese English learner corpus, and 1000 compositions were chosen. Among them, 720 were used as the training set, the training set for neural network training was somewhat smaller, and the composition of the training set was not rich. If the training set were larger, the training content would be richer and the performance of the neural network and the accuracy of scoring would be improved.

2) In the semantic analysis module, this model uses more than 20 high-scoring compositions with different topics as sample articles. If more sample compositions with different topics were added, the accuracy of semantic scoring would be greatly improved, and the comprehensive scoring accuracy of compositions would also be greatly improved.

REFERENCES

- [1] Zhou J, Liao H, Liu LX. A new basic thoracoscopic surgical skill training and assessment system using automatic scoring techniques. *Surgical Endoscopy and other Interventional Techniques*, 2021, 36(5): 3076-3086.
- [2] Varban OA, Thumma JR, Dimick JB. Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy a video-based study. *Annals of Surgery*, 2021, 273(4): 766-771.
- [3] Kim JE, Park K, Jung SY. Automatic scoring system for short descriptive answer written in Korean using lexico-semantic pattern. *Soft Computing*, 2018, 22(13): 4241-4249.
- [4] Choi WY. An application of an AI chatbot automatic pronunciation scoring system to elementary school students. *English Language Assessment*, 2021, 16 (2): 167-185.
- [5] Kim DS. Automatic scoring system for picture-based English caption writing test adopting deep learning based word-embedding. *The Linguistic Association of Korea Journal*, 2021, 29(2): 1-20.
- [6] Yuan Z. Interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm. *Journal of Intelligent & Fuzzy Systems*, 2021, 40(2): 2069-2081.
- [7] Caroline G, Schlemmer P, Rodrigues E. Evaluation of nutritional status and causes of mal-nutrition. *Archivos Latinoamericanos De Nutricion*, 2019, 69(5): 317-326.
- [8] Chen ZG, Wang YJ, Chen J. Value of distal high signal vessel sign of FLAIR sequence in the establishment of secondary collateral circulation after cerebral infarction. *Boletin De Malariologia Y Salud Ambiental*, 2019, 59(5): 86-91.
- [9] Srinivasu PN, Rao TS, Olariu I. A comparative review of optimisation techniques in segmentation of brain MR images. *Journal of Intelligent & Fuzzy Systems*, 2020, 38(5): 6031-6043.
- [10] Hulber T, Kocsis ZS, Pesznyak C. A scanning and image processing system with integrated design for automated micronucleus scoring. *International Journal of Radiation Biology*, 2020, 96(5): 628-641.
- [11] Arora V, Lahiri A, Reetz H. Phonological feature-based speech recognition system for pronunciation training in non-native language learning. *Journal of the Acoustical Society of America*, 2018, 143(1): 98-108
- [12] Fogerty D, Madorskiy R, Dubno JR. Comparing speech recognition for listeners with normal and impaired hearing: simulations for controlling differences in speech levels and spectral shape. *Journal of Speech Language and Hearing Research*, 2020, 63(12): 4289-4299.
- [13] Hong YJ, Hosung N. Evaluating score reliability of automatic English pronunciation assessment system for education. *Studies in Foreign Language Education*, 2021, 3(1): 91-104.
- [14] Bequette M, Cardiel C.L.B., Cohn S. Evaluation capacity building for informal stem education: working for success across the field. *New Directions for Evaluation*, 2019 32(161): 107-123.
- [15] Nikkonen S, Korkalainen H, Toyras J. Automatic respiratory event scoring in obstructive sleep apnea using a long short-term memory neural network. *IEEE Journal of Biomedical and Health Informatics*, 2021, 25(8): 2917-2927.
- [16] Fu SX, Gu HM, Yang B. The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. *British Journal of Educational Technology*, 2020, 51(5): 1674-1692.
- [17] Lee Y. An analysis of the influence of block-type programming language-based artificial intelligence education on the learner's attitude in artificial intelligence. *Journal of the Korean Association of Information Education*, 2019, 23(2): 189-196.
- [18] Wilhelm D, Bouarfa L, Padoy N. Artificial intelligence in visceral medicine. *Visceral Medicine*, 2020, 36(6): 471-475.
- [19] Baek S, Lee HJ, Kim H. Analysis of artificial intelligence's technology innovation and diffusion pattern: focusing on USPTO patent data. *The Journal of the Korea Contents Association*, 2020, 20 (4): 86-98.
- [20] Kim GS. A study on the trust of artificial intelligence. *Korean Journal of Local Government & Administration Studies*, 2020, 34 (3): 21-41.
- [21] Lew M,d Wilbur DC. A novel approach to integrating artificial

- intelligence into routine practice. *Cancer Cytopathology*, 2021, 129(9): 677-678.
- [22] Westcott RJ, Tchong JE. Artificial intelligence and machine learning in cardiology. *Jacc- Cardiovascular Interventions*, 2019, 12(14): 1312-1314.
- [23] Ahmed A, Agarwal S. Teaching an old dog new tricks: three-dimensional visual spatialisation of viscoelastic testing and artificial intelligence. *Anaesthesia*, 2020, 75(8): 1006-1009.
- [24] Jang C, Sung W. A study on policy acceptance intention to use artificial intelligence-based public services: focusing on the influence of individual perception & digital literacy level. *Informatization Policy*, 2022, 29 (1): 60-83.
- [25] Wilhelm D, Padoy N. Artificial Intelligence in Medicine: Passing Hype or the Holy Grail of Solutions? *Visceral Medicine*, 2020, 36 (6): 425-427.
- [26] Kim SA. Research trends in elementary and secondary school artificial intelligence education using topic modeling and problems in technology education. *The Korean Journal of Technology Education*, 2021, 21(1): 106-124.
- [27] Bhattacharya S, Pradhan KB, Singh A. Artificial intelligence enabled healthcare: A hype, hope or harm. *Journal of Family Medicine and Primary Care*, 2019, 8(11): 3461-3464.
- [28] Teramoto A. Application of artificial intelligence in radiology. *Gan to Kagaku Ryoho. Cancer & Chemotherapy*, 2019, 46(3): 418-422.
- [29] Lu HM, Li YJ, Serikawa S. Brain intelligence: go beyond artificial intelligence. *Mobile Networks & Applications*, 2018, 23(2): 368-375.
- [30] Kim T, Jung HR. A study on the utilization of artificial intelligence technology and technology management in criminal justice procedures. *Journal of Business Administration & Law*, 2020, 31(1): 581-610

An Efficient Deep Learning with Optimization Algorithm for Emotion Recognition in Social Networks

Ambika G N¹, Dr. Yeresime Suresh²

Assistant Professor, CSE Dept., BMS Institute of Technology and Management, Bangalore-560064, India¹
Associate Professor, CSE Dept., Ballari Institute of Technology & Management, Ballari-583104, India²

Abstract—Emotion recognition, or computers' ability to interpret people's emotional states, is a rapidly expanding topic with many life-improving applications. However, most image-based emotion recognition algorithms have flaws since people can disguise their emotions by changing their facial expressions. As a result, brain signals are being used to detect human emotions with increased precision. However, most proposed systems could do better because electroencephalogram (EEG) signals are challenging to classify using typical machine learning and deep learning methods. Human-computer interaction, recommendation systems, online learning, and data mining all benefit from emotion recognition in photos. However, there are challenges with removing irrelevant text aspects during emotion extraction. As a consequence, emotion prediction is inaccurate. This paper proposes Radial Basis Function Networks (RBFN) with Blue Monkey Optimization to address such challenges in human emotion recognition (BMO). The proposed RBFN-BMO detects faces on large-scale images before analyzing face landmarks to predict facial expressions for emotional acknowledgment. Patch cropping and neural networks comprise the two stages of the RBFN-BMO. Pre-processing, feature extraction, rating, and organizing are the four categories of the proposed model. In the ranking stage, appropriate features are extracted from the pre-processed information, the data are then classed, and accurate output is obtained from the classification phase. This study compares the results of the proposed RBFN-BMO algorithm to the previous state-of-the-art algorithms using publicly available datasets derived from the RBFN-BMO model. Furthermore, we demonstrated the efficacy of our framework in comparison to previous works. The results show that the projected method can progress the rate of emotion recognition on datasets of various sizes.

Keywords—Blue monkey optimization (BMO); deep learning; electroencephalograph (EEG); emotion recognition; human-computer interaction (HCI); radial basis function networks (RBFN)

I. INTRODUCTION

This template, modified in MS Word 2007 and saved as a "Word 97-2003 Document" for the PC, provides authors with most of the formatting specifications needed for preparing electronic versions of their papers. All standard paper components have been specified for three reasons: (1) ease of use when formatting individual papers, (2) automatic compliance to electronic requirements that facilitate the concurrent or later production of electronic products, and (3) conformity of style throughout conference proceedings.

Margins, columns widths, line spacing, and type styles are built-in; examples of the type styles are provided throughout this document and are identified in italic type, within parentheses, following the example. Some components such as multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.

Language, text, action, and other means are all ways that people can express themselves. How to recognize and accurately detect human facial expressions has emerged as a hot research area given the rapidly expanding artificial intelligence field [1]. Numerous businesses including amusement, security, online education, and intelligent medical care, use facial expression detection technologies [2]. Facial expression is a critical factor in human emotion recognition. Since a person's facial expressions convey their emotions, "facial recognition" [3] and "emotion recognition" are often used synonymously. Significant progress has been made in the automotive industry, augmented robotics, reality, neuromarketing, and interactive games. There is growing interests in enhancing all facets of human-computer interaction, particularly in recognizing human emotions.

Facial expression recognition can be used to monitor driver fatigue. An alarm is sent when the driver's face exhibits signs of drowsiness, and a camera records the driver's expression in real time while also analyzing the driver's mental state. This can assist with avoiding traffic accidents induced by fatigued driving. The elderly can benefit from installing a human-computer interaction system with a recognition of facial expressions feature in nursing homes or elderly homes. Facial expression recognition technology can track how each student responds to the lecture and provide the instructor with immediate feedback, which can, to some extent, advance the superiority of education [4]. During the online teaching process, it can be difficult for the instructor to keep track of each student's reaction, but it is still important to make timely adjustments to the course progress.

In the conventional method for recognizing facial expressions, a photograph is taken, its attributes are extracted, and then the image is identified using machine learning [5]. The difficult feature extraction method and the identification performance being easily influenced by the environment and a person's facial activity are some drawbacks of this strategy.

One of today's most vital and challenging techniques is emotional recognition. Applications for emotion recognition include helping to measure stress levels and blood pressure, among other things. When using emotional techniques, one can apply the functions of happy, sad, calm, and neutral facial features. The human body's inner workings can be detected using various methods and algorithms. Real-time emotional recognition can pick up on human thought processes. Identifying diseases early using emotional recognition shields can save humans from severe infections or illnesses. Emotional recognition has the main benefit of assisting in identifying human mentalities without using questions.

Machine learning algorithms accurately predict facial emotions like stress and sadness. The results for emotion recognition, such as sadness and rage, were improved when the ECG and PPG were merged with the 28 features take out from algorithms using machine learning [6]. Without knowledge sharing, facial expressions are essential for determining human mentality. In a few articles, datasets from 2010 to 2021 are combined, along with the majority of the features collected and categorized using deep learning and to support vector machine approach, hence increasing classification accuracy and outcomes.

A subset of machine learning methods called "deep learning" can be used to analyze facial expressions and identify emotions. However, the amount of data will affect how well it works. As data volume rises, performance gets better. Deep learning cannot be applied to facial expression datasets because they are too small. Several studies have found that augmentation techniques like cropping, scaling, translating, or mirroring during the pre-processing stage increase the alteration and, subsequently, the quantity of information.

In various pattern recognition and classification issues, neural networks have been used because they have the best approximation capability. Along with the back-propagation algorithm, face recognition has also used convolution neural networks and multilayer perceptron (MLPs) [7]. Because of its slow convergence rate and uncertainty about whether it will reach global optimums, the back-propagation learning procedure is computationally intensive. Due to their outstanding approximation accuracy and quick processing, radial basis function neural networks (RBFN) with a single hidden layer have been used for facial recognition applications. Radial basis functions in the hidden layer nonlinearly map the contribution face information to linearly divisible information in hidden hyperspace. Some enterprise challenges for hidden layers include defining the RBF unit centers of hidden neurons, their numbers, and the selection and shape of fundamental functions. Second, the success of blue monkey swarms naturally inspired the development of the Blue Monkey (BM) approach, a cutting-edge metaheuristic optimization method. The total number of men in a group is determined via the BM process. Like other forest guenons, blue monkey groups typically only contain one adult male outside the breeding season. With constraints and an unknown search space, this algorithm effectively finds solutions to practical problems. The BM method has some variables and the potential to produce better results [8].

Radial Basis Function Networks (RBFN) with Blue Monkey Optimization is suggested in this paper (BMO). The proposed RBFN-BMO first recognizes faces on large-scale imageries after assessing face landmarks to approximate representations for reaction acknowledgment. The two stages of the RBFN-BMO are convolutional neural networks and patch cropping. The proposed perfect is composed of four categories: feature extraction, ranking, preprocessing and organization. After preprocessing the dataset's collected data for data cleansing, the information is classified, the relevant attributes are extracted from the preprocessed data in the ranking phase, and the correct information is obtained. This study compares the results of the recommended RBFN-BMO method to earlier state-of-the-art methodologies using publicly accessible datasets inferred from the RBFN-BMO model. Furthermore, we have demonstrated that our structure is more efficient than earlier ones.

The projected work is defined in detail below. Section II goes into more excellent aspects of the work associated with the projected outcome. Section III goes over the proposed framework in depth. Section IV goes into great detail about experiment design and performance evaluation. Section V concludes the discussion of future work.

A. Contribution

- Data cleaning and pre-processing are performed on the dataset's collected data.
- The pre-processed data are given the appropriate characteristics during the ranking phase, and the information is then categorized.
- Pre-processing, feature extraction, ranking, and categorization are the four categories that make up the suggested model.
- Before looking at face landmarks to predict facial expressions for emotion detection, the Radial Basis Function Networks (RBFN) - Blue Monkey Optimization (BMO) proposed method first recognizes faces on large-scale images.

II. LITERATURE SURVEY

Chen et al. [9] used a deep sparse auto-encoder network based on Soft-max regression to identify facial expressions of emotion during human-robot interaction. This work minimizes distortion, learning efficiency is determined, and dimensional complexity is measured using the SRDSAN technique. The soft-max simple regression perfect will help categorize the input signal, while the DSAN technique helps with accurate feature extraction.

Babajee et al. [10] suggested using deep learning to recognize human expressions from facial expressions. This paper uses deep knowledge and a convolutional neural network to offer seven methods for identifying facial emotions. This study uses the Facial Action Coding System (FACS) to collect 32,398 facial expressions to identify various types of emotion. The identification approach's failure to be an effective optimization technique is the sole justification for this research.

Satyanarayana et al. [11] used deep learning and cloud access to implement emotional acknowledgment in this study. One of the most effective methods in many presentations is facial emotion recognition. Face recognition makes extensive use of the deep learning algorithm. Her thesis paper examines a variety of emotions, such as sadness, joy, serenity, and rage. The Python code generates a particular IP address for each technique to send this data.

Jayanthi et al. [12] used deep classifiers to develop an organizing strategy for emotional categorizations utilizing speech and static images. One of the most crucial methods for determining someone's stress level is emotion identification. The two traits of emotion perception and speech modulation are essential in determining the stress level in the human body.

Sati et al. [13] used NVIDIA to implement face detection, recognition, and emotion recognition in his paper. In this study by Jetson Nano, face emotion recognition and detection are combined. Facial emotional identifications have historically been among the most challenging techniques to master. This technique's accuracy and classification outcomes can be improved by adding some features. The ANN technique assists in recognizing and classifying facial expressions of emotion.

Wang et al. [14] applied a recently developed deep learning technique in this paper. The four-category deep learning model is used in this paper. Convolutional neural networks and deep architectures fall under the first category. The deep learning model has a significant impact on deep neural networks. This component of the machine learning algorithm is essential. The classification, which includes both linear and nonlinear special functions, is necessary for the accuracy of the data.

The multi-label convolution neural network was implemented by Ekundayo et al. [15] to identify facial expressions and estimate ordinal intensity. This was completed because, while many features, like FER, are consistent with the emotional recognizing method, only one is ideal for the multi-class dynamic classification method. This study utilized a multi-label convolutional neural network.

Using facial expressions, EEG, and machine learning techniques, Hassouneh et al. [16] developed a real-time emotional recognition system. The system could distinguish between smiling, remaining neutral, and losing control. In this study, virtual markers detect facial regression using the optical flow algorithm, since it acknowledges a lower level of computational complexity, the optical flow algorithm system aids in providing people with a physical challenge.

Neuro-sense was used by Tan et al. [17] to implement short-term emotion recognition and comprehension. Spiking neural network simulations of the spatiotemporal EEG patterns served as the foundation for their strategy. The SNN method is used for the first time in this paper. It aids in comprehending how the brain operates. One of the two methods used to analyze the EEG data is arousal-valence space. The various types of segments that make up the arousal-valence space include low arousal, high arousal, high

valence space techniques, and standard valence space methods.

A deep facial expression recognition survey was carried out by Li et al. [18]. One of the system's most significant challenges is recognizing a person's facial expression. The two main challenges to accurate facial expression recognition are a need for training sets and undesirable emotion variations (FER). The data set is first organized using the neural pipeline technique. Consequently, the FER technique's complex problems will be reduced.

Yang et al. [19] used a stacked auto-encoder to implement three-class profound learning-based emotional expressions. This paper demonstrates that discrete entropy calculation can be used to measure the EEG signal. The deep learning algorithm's auto-encoder technique results are more accurate than those from the encoding system's calculation methods. To use the alpha, beta, and gamma values, this method assesses emotions. Classification results are produced more accurately when a deep learning procedure is used. The deep learning procedure typically yields acceptable outcomes for the various classes of emotional recognition.

Yadahalli et al. [20] used a deep learning technique to recognize facial micro expressions. This study uses six different emotional expressions—happy, sad, angry, scared, neutral, and surprised faces—to collect the eight layers of the dataset. Because it has started collecting datasets that contain the FER perfect, the paper assumes that the FER with a CNN improves accuracy and that the removed consequence produces the multimodal facial expression using a single method. Using a single algorithm, the multimodal facial expression is also added to the dataset that has been gathered.

Asaju et al. [21] used a temporal method to recognize facial emotional expressions. This study introduces a CNN-based deep learning procedure that can implement numerous kinds of emotional acknowledgment in the human body. To extract features, VGG-19 methods are used. Both the accurate mapping technique and the recognition of facial emotions method use the BiLSTM architecture.

Yolcu et al. [22] suggested a deep learning-based method for tracking consumer performance patterns that measured head pose assessment and analyzed facial expressions to gauge the level of interest in the product. To follow customer interest, this was done. Deep structured learning was suggested by Walecki et al. [23] as a technique for determining the level of facial expression intensity. Face physics and other pre-processing methods are less critical in deep learning-based face recognition. CNN's convolution layers convolve the input image using various filters. It generates a feature map with fully connected networks to recognize facial expressions.

To identify emotional expressions on faces, Asaju et al. [24] employed the temporal method. This study integrates a CNN with a deep learning technique to create dissimilar kinds of emotional identification in the human body. Use the VGG-19 methods for obtaining features. The title and precise mapping of facial expressions are then carried out using the Bi-LSTM architecture.

Ekundayo et al. [25] implemented a multilabel convolution neural network to recognize facial expressions and estimate ordinal intensities. Although many features can be used with sentimental character segmentation techniques like FER, they are only suitable for some of them for the ideal categorization emotional classification method. Convolutional neural networks with multiple labels are used in this study.

III. PROPOSED SYSTEM

This study suggests combining RBFN and Blue Monkey Optimization (BMO) to acknowledge faces and emotions in high-resolution images. Modules for pre-processing, feature extraction, ranking, and classification can be seen in the block diagram of the proposed system in Fig. 1.

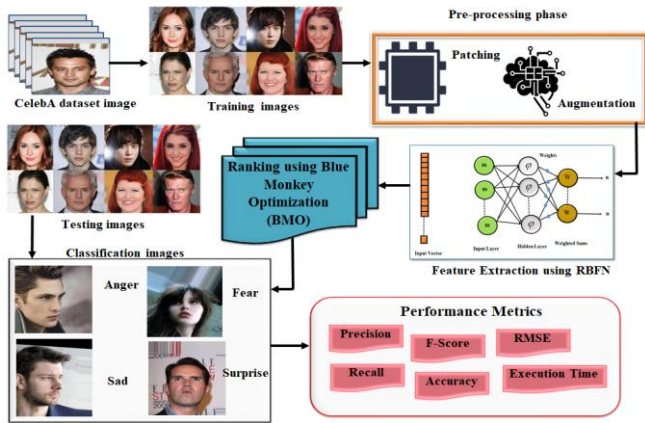


Fig. 1. Proposed method of RBFN-BMO method.

A. Pre-Processing of the Images

Before using the pyramid images approach, the original photographs are downscaled by a feature of 2 until they are 128×128 in dimensions. Additionally, each pyramid image is split into 128×128 pixel-sized patches to simplify processing and reduce memory allocation.

1) *Data augmentation*: Dawn sampling was performed on the input photos using a ratio of 2 up to an image size of 128×128. There was no overlap between the 128×128 patches created from the pyramid photos. If the picture segment is too long for the patches, zeros will be added to make up the difference. Additionally, we employ a data augmentation technique that involves rotating, translating, and flipping the images on their axes at 45-degree perspectives. The data augmentation method increases the number of training examples and progresses the network's ability to simplify under challenging circumstances. The images have been edited to resemble actual facial recognition scenarios. Due to the GPU's memory capacity, we use a batch size of 16 pictures during training. As a result, various combinations of patches for the same image might be created. The initial photo incorporates all sensing forecasts on the same image's patches.

B. Feature Extraction using RBFN

Feed forward neural networks comprise RBFN. The three layers that make up the RBFN's design are the input, hidden, and output layers, as depicted in Fig. 2. Data is sent from the

user's input layer to the concealed layer [26]. Radial basis function (RBF) units, known as hidden neurons, comprise the hidden layer. Each jth RBF unit has a related basis function (φ_j), spread ($\hat{\sigma}_j$), and centre (A_j).

The nonlinear basis functions φ_j are affected by how far the input is from the center of the jth RBF unit. RBFN frequently employs the basic functions Gaussian, multiquadric, inverse multiquadric, and thin spline. The most often used Gaussian basis function was signified in this study as

$$\varphi_j(x) = e^{-\frac{\|m-A_j\|^2}{2\hat{\sigma}_j^2}} \quad (1)$$

The restriction r, which also denotes the function's width, characterizes the spread of the radial basis function, where C_j stands for the jth RBF unit's center. You can think of RBFN as a mapping from (2).

$$R^d \rightarrow R^u \quad (u \gg d) \quad (2)$$

Where u is the amount of RBF units and $P \in R^d$ is the d-dimensional input feature vector. The ith output of the RBFNN, $n_i(m)$, is

$$n_i(m) = \sum_{j=1}^u \varphi_j(x) \times w_{i,j} \quad (3)$$

Where $w_{i,j}$ denotes the degree of connectivity between the ith output neuron and the jth RBF unit.

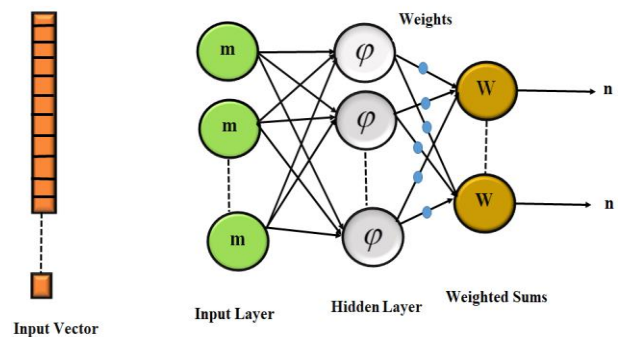


Fig. 2. Structure of RBFN.

C. Ranking with Blue Monkey Optimization (BMO)

Blue Monkeys are different from the other species. They frequently live in societies where women are the majority, meaning females stay in their natal groups. However, as soon as they reach the mature stage, the males leave their groups [27]. There are usually a lot of females and young in each group of blue monkeys, but only one male. This problem exacerbates the problem of inbreeding. The men leave the group and join another one when they're older. However, finding a new group might take some time, so the males might

initially seem to be by themselves. Regarding social interactions, blue monkeys don't have a perfect sense of intuition [28]. Social interaction only lasts for a short time, typically when playing with and grooming other people.

Babies also connect with the other adults in the group and their mothers. The source of those newborns typically avoids their male counterparts. Baby handlers are the ones who do all the work. The young females tend to the babies and carry and protect them. From this habit, infants pick up how to react like all monkeys.

1) *Group division*: The BMO algorithm imitates the actions of Blue Monkey. Each group of monkeys had to travel through the search area to simulate these interactions. Earlier, it was mentioned that when the monkeys are divided into groups, they start searching for food sources far away in an area where more robust monkeys are out of sight of conventional vision. The young *Cercopithecus mitis* and the male have little to no interaction. Because *Cercopithecus mitis* is a territorial species, young males should venture outside as soon as possible. The dominant male of another family will challenge them. If they succeed in eliminating him, they will assume control of the family and be able to provide young men with food, shelter, and socialization. Blue monkey groups typically consist of a sizable number of females and young, with only one male [29].

2) *Position update*: Each blue monkey in a group updates to the position in the best place within that group. Equations like the ones below describe this behavior:

$$Power_{m+1} = (0.7 * Power_m) + (W_{lea} - W_m) * rand * (Y_{best} - Y_m) \quad (4)$$

$$Y_{m+1} = Y_m + Power_{m+1} * rand \quad (5)$$

W_{lea} stands for the leader weight, W_m for the monkey weight, Y_{best} for the leader location, which can take any value between [0,1], $Power$ for the monkey power rate, and so on.

Using the following equations, the blue monkey's offspring are also updated.

$$Power_{m+1}^C = (0.7 * Power_m^C) + (W_{lea}^C - W_m^C) * rand * (Y_{best}^C - Y_m^C) \quad (6)$$

$$Y_{m+1}^C = Y_m^C + Power_{m+1}^C * rand \quad (7)$$

W_m^C is the child weight at which all weights are random amounts among [4, 6], Y_{best}^C is the child position, Y_{best} is the leader child position, $Rate$ needs to stand for the child power rate, is the leader child weight, and "rand" denotes an arbitrary amount among [0, 1]. Every cycle, the location needs to be reorganized.

Algorithm for Blue Monkey Optimization

```

Initialize the Blue Monkey and their children population
Bm (m=1,...,n)

Initialize power rate and weight of Blue monkey as Power
and W

Where (Power ∈ [0, 1]) (W ∈ [4, 6])

Randomly Distribute Blue Monkey into T groups and
children in one group.

Evaluate the fitness of blue monkey and children in each
group

Select Worstfit and Best fit in  $Y_{lea}^C$  Children group

T=1

While (T ≤ maximum number of group)

    Swap Worstfit =  $Y_{lea}^C$ 

    Update Power and Y position of all blue monkey by
    Eq. (4,5)

    Update Power and Y Position of children by Eq. (6,7)

    Upgrade the fitness of all blue monkeys and kids.

    Most recent Update

    If (New best > Current best) then

        New best = Current best

    End if

T=T+1

End While

The best blue monkey should be returned.
    
```

D. Classification of Emotions Recognition

The centers of hidden neurons or RBFN units can be found using the sub-clusters produced by the proposed RBFN-BMO technique. Since the method evolves the sub-clusters based on the given training statistics, providing the number of sub-clusters for each participant as input for face recognition is unnecessary. Emotion recognition aims to identify a person's emotions. The capacity to perceive emotions varies significantly among people. A significant area of research is emotion recognition with technological assistance. Analyzing facial landmarks, with a focus on the lips, nose, and eyes, is the foundation for identifying emotions. Facial expressions are a huge help in recognizing emotions. The shapes of those sections represent the person's emotions. One can infer a person's emotional state from the points' locations and the distance between them. The main objective was to create the proposed methodology for finding 68 spots; A potential area for emotion detection in each of the locations. The jawline is depicted in points 1 through 17, the left and right brows in points 18 through 22, the left and right eyes in moments 37

through 48, the noise in points 28 through 36, the outer lip area in matters 49 through 60, and the inner lip structure in points 60 through 68. In studies of facial expressions, the location of those points is crucial.

In this study, we suggest using facial features as annotations for emotion recognition. The proposed RBFN-BMO accepts inputs such as high-resolution photos, facial feature annotations, and the face's position. The seven emotions portrayed in this piece are anger, disgust, fear, joy, sadness, surprise, and neutrality. More details about the data set used to train the recommended RBFN-BMO are provided in the following section.

IV. EXPERIMENTAL RESULTS

Details about the experimental setting that was used to develop and evaluate the suggested RBFN-BMO are provided in this section. The experiment was run on a Linux desktop with an Intel i7 processor, 32 GB RAM, and a 4 GB Nvidia GTX960 graphics card. The suggested RBFN was developing using the TensorFlow deep learning framework, cuDNN, and CUDA vibration libraries. For image manipulation, the OpenCV library was used.

A. Dataset Description

We suggest using the celeb datasets [30] to train the suggested RBFN-BMO. There are 202,599 RGB images of well-known people in this dataset. Face, attribute, and landmark acknowledgment was considered when creating the dataset. The CelebA dataset's images are summarised in Fig. 3. This work advocates using facial landmarks and distinctive analysis for face detection and expression acknowledgment.

A training set, a validation set, and a testing set were created from the collected data. The training set consisted of 70% of the data, the validation set of 10% of the training set, and the testing set consisted of 30% of the data.

While pre-processing reduces filter noise, 70% of the data are used as training input. The high dimensionality of the filter is reduced with the help of feature extraction. To decrease the size of the problem space, our work employs various methods for extracting features, such as edge detection. As a result, it generates clear output images with precise dimensional quality. In the data, categorization and part extraction happen simultaneously.

B. Performance Metrics

The effectiveness of the suggested work is assessed using the recognition rate. The classification performance is calculated by dividing the total amount of images into the data sets by the number of facial expressions successfully identified. It is shown as:

The percentage of accurate classifications is measured by precision (Prec). It can be indicated using (8):

$$Precision = \frac{TP}{TP + FP} \quad (8)$$



Fig. 3. Celebrity images data set.

It is also possible to refer to the actual positive rate as the recall rate or recall. It assesses how frequently a classifier gives the right category a favorable result. It is described in (9).

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

The F-Measure represents the harmonic mean of sensitivity and precision (F). It is crucial since higher accuracy typically interprets into lower sensitivity. It can be calculated using (10).

$$F - Score = 2 * \frac{precision * recall}{precision + recall} \quad (10)$$

Accuracy: To calculate the precision of our predicted value, divide all values by the total of true negatives and true positives.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Root Mean Square Error (RMSE)

The only departure from RMSE is the square root sign. The mean absolute error equation is given in (12).

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (12)$$

The messages TP, TN, FP, and FN are confirmed positive, true negative, true positive, and false negative, respectively. The outcome should improve as a component of precision, recall, f-measure, and sensitivity.

1) *Precision analysis*: In Fig. 4 and Table I, the precision of the RBFN-BMO strategy is contrasted with that of other methods currently in use. The graph demonstrates the increased efficiency and precision of the deep learning approach. In comparison to the GRU, LSTM, RNN, DNN, and ANN models, which have precision values of 89.029%, 85.536%, 90.927%, 88.435%, and 93.983% for the 1000 data, the RBFN-BMO model has a precision value of 96.425%. With different data sizes, the RBFN-BMO model has shown its greatest performance. The RBFN-BMO's precision value is 97.927% under 6000 data, compared to the GRU, LSTM, RNN, DNN, and ANN models' precision values of 89.827%, 86.324%, 93.782%, 87.625%, and 95.029%, respectively.

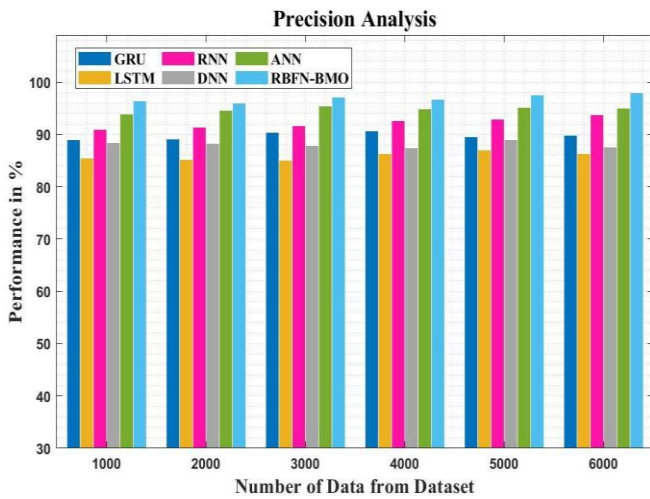


Fig. 4. Precision analysis for RBFN-BMO method with existing systems.

TABLE I. RBFN-BMO METHOD PRECISION ANALYSIS USING EXISTING SYSTEMS

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	89.029	85.536	90.927	88.435	93.983	96.425
2000	89.214	85.234	91.425	88.323	94.626	96.029
3000	90.425	85.029	91.728	87.922	95.435	97.182
4000	90.627	86.324	92.637	87.425	94.928	96.728
5000	89.526	86.973	92.938	88.937	95.227	97.632
6000	89.827	86.324	93.782	87.625	95.029	97.927

2) *Recall analysis*: Fig. 5 and Table II illustrate how the RBFN-BMO approach compares to other current methods in terms of recall. The figure demonstrates how the recall performance was enhanced by the deep learning approach. The RBFN-BMO model, for example, has a recall value of 93.827% with 1000 data, while the GRU, LSTM, RNN, DNN, and ANN models have recall values of 79.637%, 80.928%, 84.938%, 86.927%, and 89.627%, respectively.

and ANN models have recall values of 79.637%, 80.928%, 84.938%, 86.927%, and 89.627%, respectively. However, the RBFN-BMO model worked most effectively with various data sizes. For 6000 data points, the recall value of the RBFN-BMO is 95.737% as opposed to the GRU, LSTM, RNN, DNN, and ANN models' respective recall values of 81.924%, 84.536%, 87.736%, 90.326%, and 92.413%.

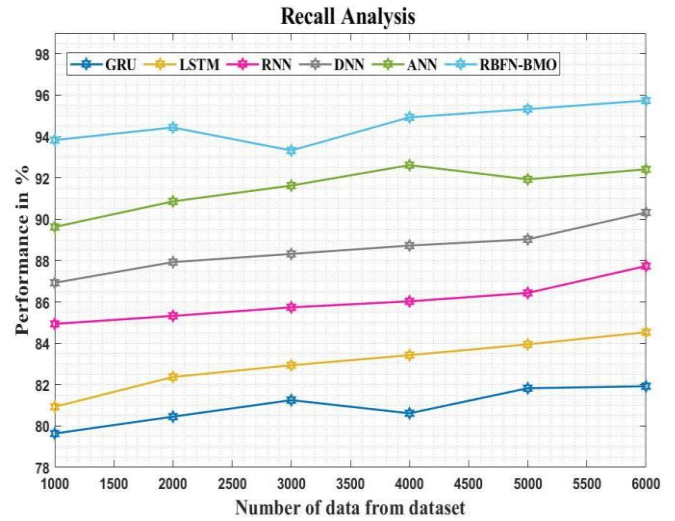


Fig. 5. Recall analysis for the RBFN-BMO method with existing systems.

TABLE II. RECALL ANALYSIS FOR THE RBFN-BMO METHOD USING EXISTING SYSTEMS

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	79.637	80.928	84.938	86.927	89.627	93.827
2000	80.452	82.373	85.324	87.926	90.862	94.435
3000	81.252	82.938	85.738	88.322	91.627	93.324
4000	80.615	83.425	86.029	88.726	92.617	94.928
5000	81.827	83.948	86.435	89.028	91.928	95.324
6000	81.924	84.536	87.736	90.326	92.413	95.737

3) *F-Score analysis*: Fig. 6 and Table III display an f-score contrast of the RBFN-BMO strategy with other existing methods. The graph shows that the deep learning method has produced better performance regarding the f-score. The RBFN-BMO model, for example, has an f-score of 92.536% with 1000 data, while the GRU, LSTM, RNN, DNN, and ANN models have f-scores of 87.928%, 85.435%, 81.526%, 83.425%, and 90.324%, respectively. The RBFN-BMO model, on the other hand, has performed best over a range of data sizes. Similarly, the f-score value of RBFN-BMO under 6000 data is 94.627%, while for GRU, LSTM, RNN, DNN, and ANN models, it is 89.928%, 87.435%, 82.213%, 85.324%, and 91.928%, respectively.

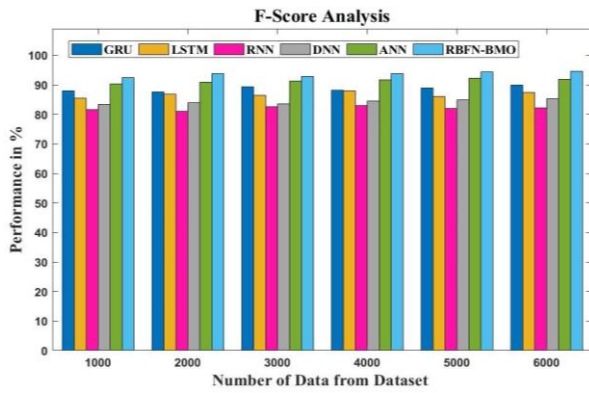


Fig. 6. F-Score analysis for the RBFN-BMO method using existing systems.

TABLE III. F-SCORE ANALYSIS USING TRADITIONAL SYSTEMS USING THE RBFN-BMO METHOD

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	87.928	85.435	81.526	83.425	90.324	92.536
2000	87.526	86.928	81.029	83.928	90.928	93.727
3000	89.432	86.425	82.536	83.526	91.243	92.927
4000	88.214	87.928	82.938	84.525	91.627	93.826
5000	88.921	86.029	81.937	84.928	92.173	94.324
6000	89.928	87.435	82.213	85.324	91.928	94.627

4) *Accuracy analysis*: Fig. 7 and Table IV shows the accuracy of the RBFN-BMO approach as compared to that of other currently utilized approaches. The graph shows how deep learning improves performance with accuracy. The RBFN-BMO model, for example, has a 1000-data accuracy of 97.627%, whereas the GRU, LSTM, RNN, DNN, and ANN models have accuracy of 89.536%, 90.917%, 92.524%, 94.526%, and 96.425%, respectively. The RBFN-BMO model, on the other hand, fared well with varying data sizes. Similarly, the accuracy of the RBFN-BMO under 6000 data is 99.546%, while the accuracy of the respective GRU, LSTM, RNN, DNN, and ANN models is 90.716%, 92.817%, 93.926%, 95.736%, and 97.125%.

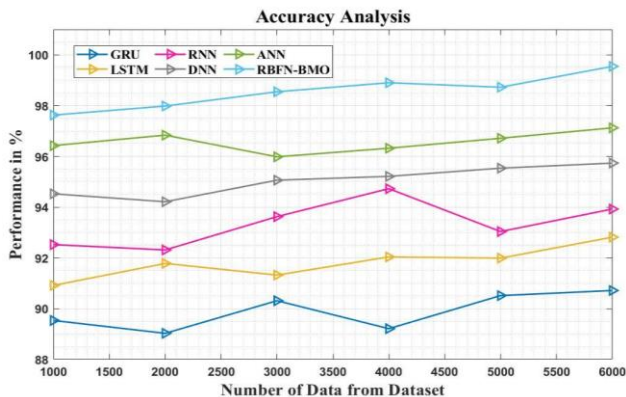


Fig. 7. Accuracy analysis for RBFN-BMO method with existing systems.

TABLE IV. ANALYSIS OF RBFN-BMO METHOD ACCURACY WITH EXISTING SYSTEMS

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	89.536	90.917	92.524	94.526	96.425	97.627
2000	89.029	91.783	92.314	94.213	96.837	97.983
3000	90.314	91.324	93.627	95.063	95.983	98.546
4000	89.213	92.039	94.728	95.213	96.322	98.902
5000	90.516	91.992	93.039	95.536	96.714	98.724
6000	90.716	92.817	93.926	95.736	97.125	99.546

5) *RMSE analysis*: Fig. 8 and Table V display an RMSE similarity between the RBFN-BMO strategy and other earlier techniques. The graph shows that the deep learning strategy has produced better results with a lower RMSE value. For instance, the RMSE value for the RBFN-BMO is 25.637% with 100 data, while the RMSE values for the GRU, LSTM, RNN, DNN, and ANN models are slightly higher at 30.526%, 26.928%, 33.626%, 36.536%, and 43.737%, respectively. The RBFN-BMO model, on the other hand, has demonstrated that it performs best with diverse data sizes while keeping a low RMSE. The RMSE value for the RBFN-BMO model under 6000 data is 26.324%, whereas it is 32.933%, 29.322%, 35.342%, 42.627%, and 47.326% for the GRU, LSTM, RNN, DNN, and ANN models, respectively.

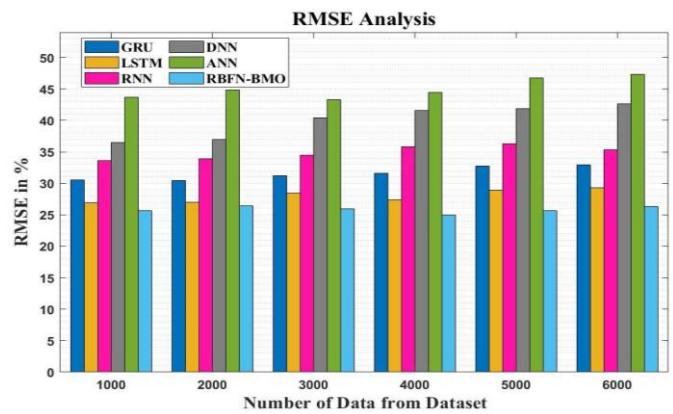


Fig. 8. RMSE analysis of the RBFN-BMO method with existing systems.

TABLE V. RMSE ANALYSIS OF THE RBFN-BMO METHOD USING EXISTING SYSTEMS

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	30.526	26.928	33.626	36.536	43.737	25.637
2000	30.425	27.029	33.928	36.928	44.863	26.435
3000	31.252	28.425	34.526	40.425	43.324	25.926
4000	31.627	27.425	35.827	41.627	44.432	25.029
5000	32.737	28.926	36.324	41.911	46.732	25.627
6000	32.933	29.322	35.342	42.627	47.326	26.324

6) *Execution time analysis*: The execution time analysis of the RBFN-BMO technique using existing methods is described in Table VI and Fig. 9. The information clearly shows that the RBFN-BMO method has outperformed the other techniques in every way. The RBFN-BMO process, for example, took only 2.738ms to execute 1000 data, while GRU, LSTM, RNN, DNN, and ANN took 12.837ms, 10.637ms, 8.526ms, 6.938ms, and 4.837ms, respectively. Similarly, the RBFN-BMO method takes 3.624ms to execute 6000 data, whereas the other existing techniques such as GRU, LSTM, RNN, DNN, and ANN have taken 15.526ms, 11.638ms, 9.553ms, 7.425ms, and 5.029ms, respectively.

TABLE VI. ANALYSIS OF RBFN-BMO METHOD EXECUTION TIME WITH EXISTING SYSTEMS

Data from dataset	GRU	LSTM	RNN	DNN	ANN	RBFN-BMO
1000	12.837	10.637	8.526	6.938	4.837	2.738
2000	12.536	10.827	8.928	6.324	4.213	2.324
3000	13.627	10.324	8.435	6.022	4.039	3.029
4000	13.829	11.526	9.073	7.425	4.637	3.927
5000	13.425	11.627	9.224	7.829	5.324	3.526
6000	15.526	11.638	9.553	7.425	5.029	3.624

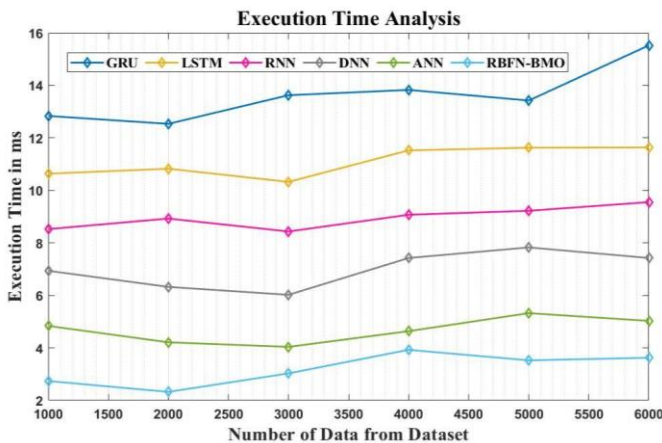


Fig. 9. Execution time analysis for the RBFN-BMO method with existing systems.

V. CONCLUSION

Facial expression analysis is one tool used to develop emotional intelligence, which is becoming increasingly important in many fields, including business and education. Numerous factors, such as visual contexts, point-of-view shifts, intra- and inter-class differences and more, impede the development of a reliable emotion recognition system. This paper suggests a convolutional neural network as a potential solution to the emotion recognition problem. To examine facial expressions in images, the proposed RBFN was made face-sensitive. The proposed RBFN-BMO can recognize people in high-resolution photos, evaluate facial expressions using facial features, and predict emotional states. The

recognition during testing validates the proposed efficacy. Regarding output quality, the established RBFN-BMO classification algorithms perform better than those currently used. The RBFN-BMO uses every dataset used for the analysis and achieves the highest level of accuracy possible, which is 99.54%. To evaluate the effectiveness of the proposed classifier, the performance of proposed RBFN-BMO is compared with the existing Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Deep Neural Network (CNN), Artificial Neural Network (ANN). As a result, the RBFN-BMO can produce better results for celebs datasets. Furthermore, it can be inferred that the Blue Monkey Optimization (BMO) meta-heuristic algorithm selects the input data features that are both the most informative and the most pertinent. It helps to achieve better categorization and reduces the error brought on by the Root Mean Square. The future of our research depends on adding new features, classifying facial emotions into their ten subcategories, and researching automatic facial emotion recognition.

REFERENCES

- [1] M. Mehdi, R. B. Vistro, E. A. Mahmoud, and H. O. Elansary, "Application of Drone Surveillance for Advance Agriculture Monitoring by Android Application Using Convolution Neural Network," *Agronomy*, 13(7), p.1764, 2023.
- [2] G. Min, G. Yukun, and T. T. Hormel, "A deep learning network for classifying arteries and veins in montaged widefield OCT angiograms," *Science*, vol. 2, no. 2, article 100149, 2022.
- [3] S. Tian Yingjie, and L. S. Duo, "Recent advances on loss functions in deep learning for computer vision," *Neurocomputing*, vol. 497, pp. 129–158, 2022.
- [4] V. Andrea, J. Sumit, and H. Shayne, "Face detection and grimace scale prediction of white furred mice," *Machine Learning with Applications*, vol. 8, article 100312, 2022.
- [5] H. Zeng, B. Zhang, B. Song et al, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018.
- [6] B. J. Park, C. Yoon, E. H. Jang, and D. H. Kim, "Physiological signals and recognition of negative emotions," in *Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 1074–1076, IEEE, Jeju, Korea, October 2017
- [7] L. Wiskott, and C. Von Der Malsburg, "Recognizing faces by dynamic link matching," *NeuroImage*, vol. 4, no. 3, pp. S14–S18, 1996.
- [8] Mahmood, Maha, and Belal Al-Khateeb. "The blue monkey: A new nature inspired metaheuristic optimization algorithm." *Periodicals of Engineering and Natural Sciences* 7.3 (2019): 1054-1066.
- [9] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," *Information Sciences*, vol. 428, pp. 49–61, 2018.
- [10] P. Babajee, G. Suddul, S. Armoogum, and R. Foogooa, "Identifying human emotions from facial expressions with deep learning," in *Proceedings of the 2020 Zooming Innovation in Consumer Technologies Conference (ZINC)*, pp. 36–39, IEEE, Novi Sad, Serbia, May 2020.
- [11] P. Satyanarayana, D. P. Vardhan, R. Tejaswi, and S. V. P. Kumar, "Emotion recognition by deep learning and cloud access," in *Proceedings of the 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 360–365, IEEE, Greater Noida, India, December 2021.
- [12] K. Jayanthi, and S. Mohan, "An integrated framework for emotion recognition using speech and static images with deep classifier fusion approach," *International Journal of Information Technology*, pp. 1–11, 2022.

- [13] V. Sati, S. M. Sanchez, N. Shoebibi, A. Arora, and J. M. Corchado, "Face detection and recognition, face emotion recognition through NVIDIA Jetson Nano," in Proceedings of the International Symposium on Ambient Intelligence, pp. 177–185, Springer, Cham, September 2020.
- [14] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," International Journal of Machine Learning and Cybernetics, vol. 11, no. 4, pp. 747–750, 2020.
- [15] O. Ekundayo, and S. Viriri, "Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation," PeerJ Computer Science, vol. 7, p. e736, 2021.
- [16] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," Informatics in Medicine Unlocked, vol. 20, Article ID 100372, 2020.
- [17] C. Tan, M. Sarlija, and N. Kasabov, "NeuroSense: short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns," Neurocomputing, vol. 434, pp. 137–148, 2021.
- [18] S. Li, and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 7, no. 3, pp. 1195–1215, 2020.
- [19] B. Yang, X. Han, and J. Tang, "Tree class emotions recognition based on deep learning using stacked autoencoder," in Proceedings of the 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMED), pp. 1–5, IEEE, Shanghai, China, October 2017.
- [20] S. S. Yadahalli, S. Rege, and S. Kulkarni, "Facial micro expression detection using deep learning architecture," in Proceedings of the 2020 International Conference on Smart Electronics and Communication (ICOSEC), pp. 167–171, IEEE, Trichy, India, September 2020.
- [21] C. Asaju and H. Vadapalli, "A temporal approach to facial emotion expression recognition," in Proceedings of the Southern African Conference for Artificial Intelligence Research, pp. 274–286, Springer, Cham, January 2021.
- [22] . Yolcu, I. Oztel, S. Kazan et al (2020) Deep learning-based face analysis system for monitoring customer interest. J Ambient Intell Human Comput 11:237–248. <https://doi.org/10.1007/s12652-019-01310-5>
- [23] R.R Walecki "Deep structured learning for facial expression intensity estimation", Image Vis Comput, vol. 259, pp. 143–154, 2017.
- [24] C. Asaju, and H. Vadapalli, "A temporal approach to facial emotion expression recognition," in Proceedings of the Southern African Conference for Artificial Intelligence Research, pp. 274–286, Springer, Cham, January 2021.
- [25] O. Ekundayo and S. Viriri, "Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation," PeerJ Computer Science, vol. 7, p. e736, 2021.
- [26] V. Agarwal, and S. Bhanot, "Radial basis function neural network-based face recognition using firefly algorithm," Neural Computing and Applications, vol. 30, no. 8, pp.2643-2660, 2018.
- [27] E. Kibebew and K. Abie "Population Status, Group Size, And Threat to Boutourlini's Blue Monkeys (Cercopithecus Mitis Boutourlinii) In Jibat Forest," Ethiopia. J Ecosyst Ecography vol. 7, pp. 230. Doi:10.4172/2157-7625.1000230.,2017.
- [28] K. Abie, A. Bekele A "Population Estimate, Group Size and Age Structure of the Gelada Baboon (The Ropithecus Gelada) Around Debre-Libanos," Northwest Shewa Zone, Ethiopia. Glob J Sci Fr R Biol Sci vol. 17, pp. 27-33, 2017.
- [29] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A Bio-Inspired Optimizer for Engineering Design Problems," Adv. Eng. Softw., Vol. 114, pp. 163–191, 2017.
- [30] Liu Z, Luo P, Wang X, Tang X (2018) Large-scale celebfaces attributes (celeba) dataset. Retrieved August 15: 2018

Improved Drosophila Visual Neural Network Application in Vehicle Target Tracking and Collision Warning

Jianyi Wu

School of Traffic Management and Engineering, Guangxi Police College, Nanning 530023, China

Abstract—To enable the vehicle tracking and collision warning system to face more complex road information, the Drosophila visual neural network collision warning algorithm has been improved, including image stabilization algorithm, target region synthesis algorithm, and target tracking algorithm. The results showed that the improved image stabilization algorithm had significantly higher image stabilization quality. The peak signal-to-noise ratio of the stabilized image before improvement was the highest at 80dB and the lowest at 54dB. After improvement, the peak signal-to-noise ratio of the stabilized image was the highest at 82dB and the lowest at 60dB. The improved algorithm did not have any false alarms or missed alarms in collision warning. In video 1, there were false alarms in the unimproved algorithm, while in video 2, there were missed alarms. In video 1, all frames were in a safe state, but the original algorithm displayed an alarm in frames 7-12, 13-22, and 23-31. In video 2, there were dangerous situations in frames 8-24 that required an alarm, while the original algorithm displayed an alarm message in frames 8-17, consistent with the actual situation. The improved target tracking algorithm can complete the task of extracting target motion curves. The target tracking algorithm extracted the motion curves of one target in video 1 and two targets in video 2, which were consistent with the video content. The improvement of the Drosophila visual neural network collision warning model through research is effective, which can improve the driving safety of vehicles in complex road conditions.

Keywords—Drosophila visual neural network; collision warning; target calibration; target tracking

I. INTRODUCTION

With the continuous development of the social economy and the continuous improvement of people's travel standards, the number of private cars has reached 402 million. However, with the continuous growth of the number of motor vehicles, the incidence of traffic accidents is also increasing, with traffic accidents accounting for over 80% of all safety accidents [1]. The occurrence of traffic accidents not only causes huge economic losses, but also seriously threatens people's life safety. In large-scale traffic accidents, the survival rate of personnel is basically 0. The lack of concentration among drivers is the main cause of traffic accidents. If a warning system can be designed to provide early warning for drivers, increase their reaction time, and prevent accidents before they occur, it can fundamentally solve traffic safety accidents. This not only protects personal safety, but also avoids economic losses. Therefore, designing an active safety system for

vehicles has become the main research direction of scholars [2]. The driver's information perception mainly relies on visual perception, but human vision has limitations. Some insects, if flies have comprehensive vision, can obtain more driving information. Therefore, we propose to use an improved fruit fly visual neural network for vehicle target tracking and collision warning, providing drivers with more driving information and ensuring safety beside the vehicle.

The study first proposed and applied an improved Drosophila visual neural network for vehicle target tracking and collision warning. This new application fully utilizes the comprehensive view of fruit flies, providing drivers with more driving information, thereby significantly improving the driving safety of the vehicle. And the innovative integrated early warning function can increase the driver's reaction time and even prevent accidents before they occur. This feature will directly reduce the occurrence of large-scale traffic accidents, protect personal safety, and avoid economic losses.

The first part is a review of the current research status of foreign vehicle collision warning and Drosophila visual neural network. The second part is the application research of improved Drosophila visual neural network in vehicle tracking and collision warning. The third part is the analysis of experimental simulation results of model application. The fourth part summarizes the research content and points out the shortcomings, and clarifies the future development direction.

II. RELATED WORKS

Effective vehicle tracking and collision warning system can help drivers avoid many traffic safety accidents. Sanberg et al. believed that the current collision warning system was based on radar system and monocular vision, with more redundancy. To reduce the redundancy of the system, a collision warning system based on stereo vision was proposed, which can detect obstacles on the vehicle path without relying on Semantic information. The final evaluation results indicated that when the obstacle was higher than 0.4m, all obstacles in the dataset can be detected [3]. To solve the anti-collision problem in autonomous vehicle, Hu et al. developed a path planning and tracking framework based on model predictive control. This framework not only considered the friction coefficient between the tire and the road, but also considered the lateral distance control of the vehicle and the adaptive weight of the speed output in various situations. The anti-collision experiment results showed that the framework developed by the author had practical significance. It can cope

with the anti-collision task of the current auto drive system [4]. To reduce the collision probability of autonomous vehicles, Cao and Jiang designed a trajectory planning and tracking control method for formation driving. The trajectory planning of this method was divided into two stages: stable speed tracking and parameter matching. The results showed that this method can effectively avoid static and dynamic obstacles, and compared with traditional controllers, this method had higher stability and more accurate tracking [5]. Zhou et al. designed a trajectory planning and tracking control strategy in order to achieve the safe driving of the driverless vehicle to the destination. This strategy uses the artificial fish school algorithm to plan the optimal path from the starting point to the end point, and uses the forward search algorithm based on Markov chain to plan the path in the local path with obstacles. The simulation results show that, the trajectory planning and control strategy proposed by the author is sufficient to face static obstacles and some dynamic obstacles [6].

Tokuda proposed a visual servo scheme based on convolutional neural networks to make the positioning of the robotic arm more accurate. In this scheme, the eye and hand cameras were used to capture the desired image and the current image to estimate the relative pose between the desired end effector and the current end effector. Simulation results showed the effectiveness of this method [7]. Burguera et al. proposed an architecture based on an automatic encoder for the rapid and stable visual loop detection of underwater robots. The decoder part of this architecture was replaced by a fully connected layer. The results showed that this neural network can improve the visual loop detection efficiency of underwater robots [8]. Gu et al. proposed a novel convolutional neural network architecture for visual availability detection in the fields of robotics and computer vision. The architecture adopted an encoder decoder architecture for acute pixel level availability detection, where the encoder network included residual modules and multi-level dependent attention mechanisms. The experimental results showed that this method improved the performance of the neural network in the attention mechanism and sampling layer networks, laying the foundation for the research on multi task learning of physical robots [9]. To solve the problem of collision perception between robots and autonomous vehicle, Q Fu et al. constructed a visual neural network based on a

lobular giant motion detector. The construction of the network referred to the visual path of locusts. In the simulation experiment, the method passed the system test of real scene stimulation, indicating that the model can effectively detect hidden obstacles under various dynamic and chaotic backgrounds [10].

In summary, the main safety issue faced by autonomous vehicle systems is the collision problem during vehicle operation. Efficient and accurate collision warning can effectively improve the safety of vehicle operation. Collision warning is based on visual detection of obstacles to avoid, and visual neural networks can effectively detect obstacle information in images to achieve vehicle collision warning.

III. VEHICLE TRACKING COLLISION WARNING BASED ON IMPROVED DROSOPHILA VISUAL NEURAL NETWORK

The main content of this chapter is to design and construct a vehicle tracking collision warning model and a target tracking and vehicle collision detection model, which is divided into two sections. The first section is the establishment of a collision detection model based on an improved Drosophila visual neural network, and the second section is the construction of a target tracking and vehicle collision warning model.

A. Collision Detection Model Based on Improved Drosophila Visual Neural Network

During the driving process of vehicles, due to obstacles or unevenness on the road surface, there may be shaking and other phenomena, resulting in missing or unreadable information recorded by the onboard camera. Therefore, the study proposes to divide the image area equally, improve the grayscale projection image stabilization algorithm, and use the improved algorithm to complete the distorted image or video [11-12]. The grayscale projection algorithm can use the grayscale value of the reference frame to complete the motion vector of the current frame, obtain a stable image sequence, study dividing the image into sub grids of equal size, and then use the grayscale algorithm to calculate the motion vector of the target within each sub grid. After integrating the motion vectors within each sub grid, the overall motion vector of the image can be obtained. The overall motion vector estimation algorithm process is shown in Fig. 1.

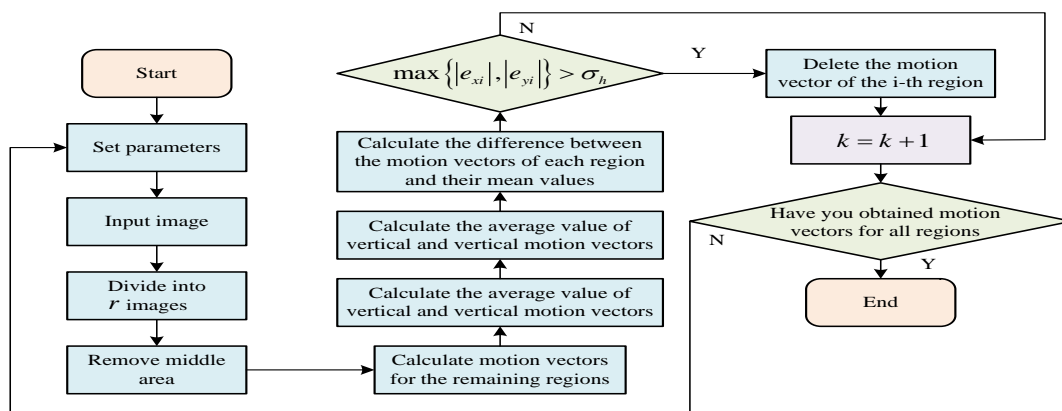


Fig. 1. Global motion vector estimation algorithm.

The threshold of the algorithm is set to σ_h , $k=1$. The image is input into the algorithm. The lower 75% of the image is divided into r small cells of the same size, the motion vectors of the small cells are calculated except for the middle region, and the set of motion vectors is obtained in the horizontal and vertical directions of the region block, as shown in formula (1).

$$\begin{cases} V_x = \{v_{x1}, v_{x1}, \dots, v_{x,r-1}\} \\ V_y = \{v_{y1}, v_{y1}, \dots, v_{y,r-1}\} \end{cases} \quad (1)$$

In formula (1), V_x, V_y represent the sum of motion vectors in the horizontal and vertical directions, and v_{x1}, v_{y1} represent the motion vectors in the horizontal and vertical directions of region 1. After obtaining the motion vectors in the horizontal and vertical directions for all regions except the central region, the average value v_{ax}, v_{ay} of the motion vectors is calculated in the horizontal and vertical directions, and then calculate the deviation value e_{xi} between v_x, v_y and v_{ax}, v_{ay} for each region, to determine whether the deviation value meets formula (2).

$$\max \{|e_{xi}|, |e_{yi}|\} > \sigma_h \quad (2)$$

If satisfied, the motion vectors of each region i in V_x, V_y are deleted, where V_x, V_y is the global motion vector of the image, $k=k+1$. The global motion vector is calculated for the next frame until the global motion vector for each frame is obtained. The above algorithm can obtain the superposition of motion vectors generated by normal shooting or shaking shooting of car mounted cameras [13-14]. The purpose of the image stabilization algorithm is to preserve the motion information obtained by the camera during normal shooting and eliminate information loss caused by shaking shooting. Therefore, a new image stabilization algorithm has been studied and designed, and its process is shown in Fig. 2.

After determining the memory scale U , maximum offset T_1 , cumulative offset $M=0$, and other parameters of the

algorithm, the stable image is output. Then the global motion vector estimation algorithm is used to calculate the global motion vector v_{xi}, v_{yi} of frame i relative to frame $i-1$, $k=k+1$. The grayscale of frame k is input, the global motion vector v_{xk}, v_{yk} of frame k relative to frame $k-1$ is calculated, and then the average motion vector $\bar{v}_{xk}, \bar{v}_{yk}$ in the horizontal and vertical directions of the adjacent L -frame grayscale before frame k is calculated. The cumulative offset M of two frames of images is calculated using formula (3).

$$M = M + \sqrt{\Delta v_{xk}^2 + \Delta v_{yk}^2} \quad (3)$$

In formula (3), $\Delta v_{xk}, \Delta v_{yk}$ represents the difference between the motion vectors of two grayscale images. At this point, if $M < T_1$, the motion vector Δ_{xk}, Δ_{yk} will be compensated for the corresponding distance in the opposite direction, and the image in frame k will be compensated and replaced. Then, $k=k+1$ will start a new round of calculation until $M \geq T_1$, and if $M \geq T_1$, the motion vector $[\Delta_{xk}, \Delta_{yk}]$ will be compensated for the corresponding pixel distance in the opposite direction to compensate for the image in frame k . Let $k=k+1$, $M=0$ start motion compensation for the next grayscale image again until all images have completed stabilization. In addition to the stability of moving images, collision detection algorithms also need to consider the influence of weather factors. Traditional weather recognition algorithms are difficult to meet the needs of weather recognition in complex environments. Therefore, research has proposed adding parameters to improve the weather recognition algorithm. Firstly, the study has added standard deviation parameters, which can reflect the degree of image dispersion. The larger the standard deviation, the smaller the impact on image target extraction, the calculation of standard deviation is shown in formula (4).

$$\sigma = \sqrt{\sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)} \quad (4)$$

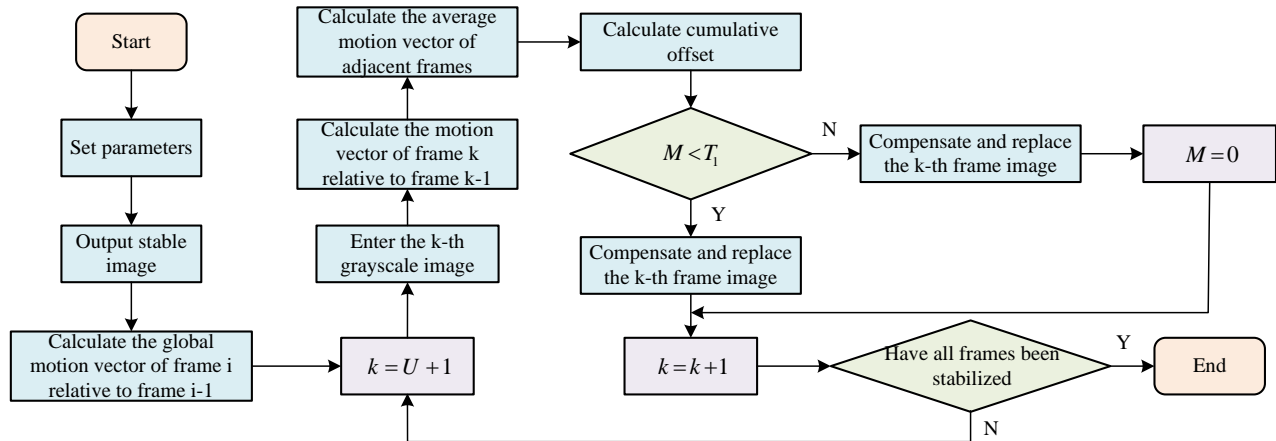


Fig. 2. Image stabilization algorithm flow.

In formula (4), σ represents the standard deviation, L represents the total number of grayscale levels, z_i represents the i -th grayscale level, $p(z_i)$ represents the probability of grayscale being z_i in the grayscale distribution of the histogram, and m is the pixel mean. Secondly, the study adds a smoothness parameter, which is an important characteristic of image texture. The higher the smoothness, the worse is the image smoothness. The calculation of smoothness is shown in formula (5).

$$R = \frac{1}{1 + \sigma^2} \tag{5}$$

In formula (5), R represents smoothness. Finally, the study also adds image entropy, which can reflect the degree of similarity between different images. The calculation of image entropy is shown in formula (6).

$$e = -\sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i) \tag{6}$$

In formula (6), e represents image entropy. After introducing the above three parameters, the indicators for measuring sunny days are shown in formula (7).

$$w_1^i = a_{10}B + a_{11}A + a_{12}C + a_{13}p_r + a_{14}\sigma + a_{15}R + a_{16}e \tag{7}$$

In formula (7), w_1^i represents the sunny day metric, a_{ij} represents the weighting coefficient, A represents the image sharpness, B represents the image brightness, C represents the image contrast, and p_r represents the proportion of image value neighborhoods. The formula for measuring indicators on cloudy days is shown in formula (8).

$$w_2^i = a_{20}(1-B) + a_{21}A + a_{22}C + a_{23}(1-p_r) + a_{24}\sigma + a_{25}R + a_{26}e \tag{8}$$

In formula (8), w_2^i represents that in addition to cloudy and sunny days, the measurement index for cloudy days also needs to consider foggy days. The measurement index formula for foggy days is shown in formula (9).

$$w_3^i = a_{30}(1-B) + a_{31}(1-A) + a_{32}(1-C) + a_{33}p_r + a_{34}(1-\sigma) + a_{35}R + a_{36}e \tag{9}$$

In formula (9), w_3^i represents the measurement indicator for foggy weather. By replacing the indicators in traditional weather recognition algorithms with the above three measurement indicators, an improved weather recognition algorithm can be obtained. The Drosophila visual neural network can detect moving targets, but it is limited to the direction detection of moving targets. After scholars' improvement, the Drosophila visual neural network can perform collision detection of moving targets. The photosensitive cell layer of the visual neural network that can perform collision detection has $I \times J$ photosensitive cells, corresponding to the grayscale value of the grayscale image of size $I \times J$, so that $L_{ij}(t)$ represents (i, j) grayscale value at time t . The output of photosensitive cell (i, j) at time t can be expressed using formula (10).

$$P_{ij}(t) = \frac{1}{2}(L_{ij} + f(L_{ij})), 1 \leq i \leq I, 1 \leq j \leq J \tag{10}$$

In formula (10), $P_{ij}(t)$ represents the output of the photosensitive cell at time t , and $f(\square)$ represents the time delay function. After combining the improved image processing technology with a visual neural network with collision detection function, an improved Drosophila visual neural network algorithm can be obtained. The algorithm process is shown in Fig. 3.

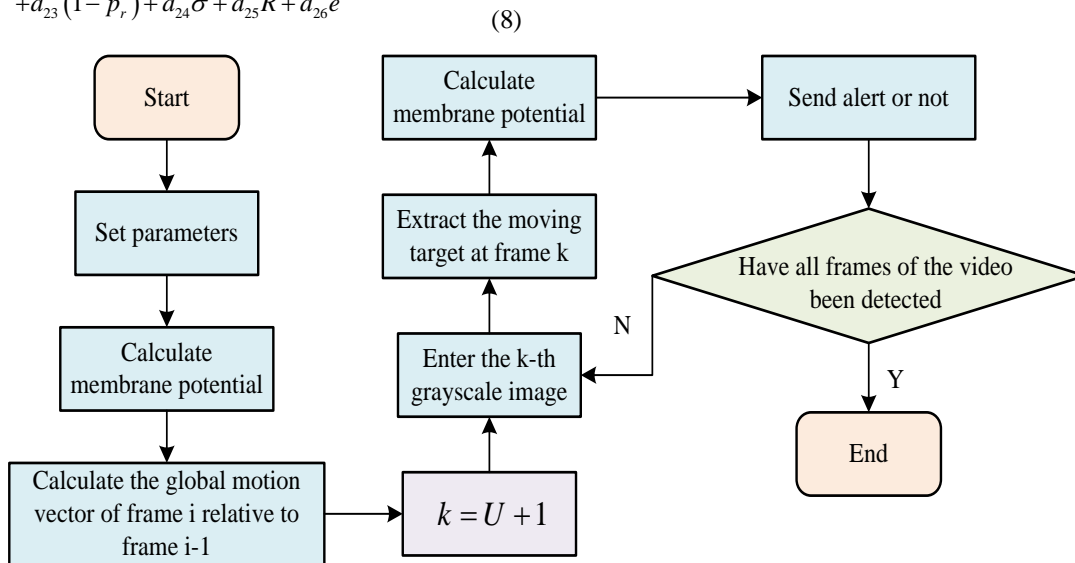


Fig. 3. Improved Drosophila visual neural network collision detection algorithm.

Initial parameters are input, neural network is used to obtain membrane potential $Y_{lob}(0), \dots, Y_{lob}(U)$, and global motion vector is calculated, $k = U + 1$. Grayscale image of frame k is input to extract foreground moving targets and neural network is used to obtain membrane potential $Y_{lob}(k)$.

If $Y_{lob}(k)$ is greater than collision warning threshold, warning signal will be sent, $k \leftarrow k + 1$. Foreground moving targets are extracted in the next frame and membrane potential. Whether warning is needed is determined until each frame of the video completes detection [15-16].

B. Construction of Target Tracking and Vehicle Collision Warning Model

In the binary image of a video, there may be a phenomenon where moving targets are divided into multiple interconnected regions, which affects the recognition of the number of moving targets in the field of view and leads to target tracking failure. Therefore, the study first uses corrosion and dilation algorithms to process the binary image, then calibrates the moving targets in the binary image, records the center point position of the targets, and finally uses region synthesis algorithms to determine the number of moving targets in the image. In a binary graph, different connected regions of the same moving target can be merged using the distance between the bounding rectangles of the moving target area and the distance threshold. The distance between the bounding rectangles is calculated using formula (11).

$$d(L_a, L_b) = \sqrt{d_h^2 + d_v^2} \tag{11}$$

In formula (11), $d(L_a, L_b)$ represents the distance between region a and region b , and d_h, d_v represent the horizontal width distance and vertical height distance of region a, b . The calculation of d_h is shown in formula (12).

$$d_h = \begin{cases} 0, & d_h^{ab} \leq \frac{1}{2}(h_a + h_b) \\ d_h^{ab} - \frac{1}{2}(h_a + h_b), & d_h^{ab} > \frac{1}{2}(h_a + h_b) \end{cases} \tag{12}$$

In formula (12), d_h^{ab} represents the horizontal projection distance between the center points of the bounding rectangle in the target area, and h_a, h_b represent the horizontal height of the target area. The calculation of d_v is shown in formula (13).

$$d_v = \begin{cases} 0, & d_v^{ab} \leq \frac{1}{2}(v_a + v_b) \\ d_v^{ab} - \frac{1}{2}(v_a + v_b), & d_v^{ab} > \frac{1}{2}(v_a + v_b) \end{cases} \tag{13}$$

In formula (13), d_v^{ab} represents the vertical projection distance between the center points of the bounding rectangle in the target area, and v_a, v_b represents the vertical distance of the target area. The target region synthesis algorithm designed for research is shown in Fig. 4.

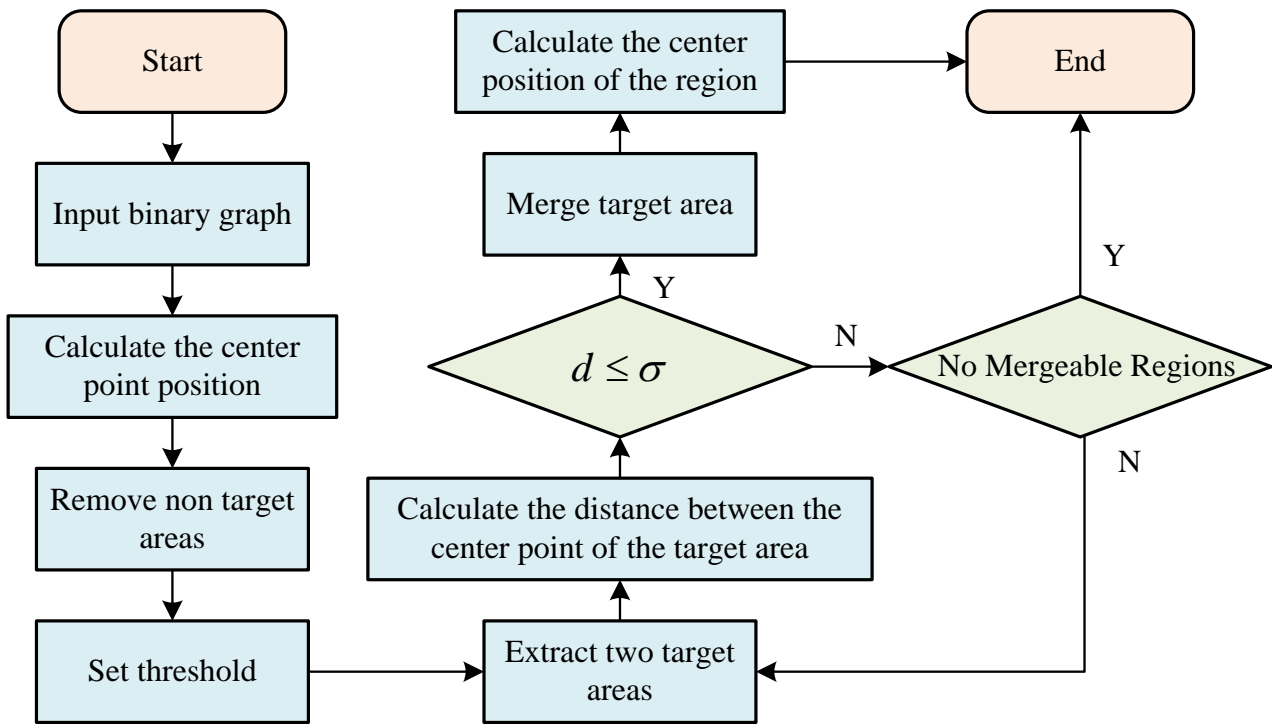


Fig. 4. Target region synthesis algorithm flow.

After inputting the binary image, the synthesized moving target in the target area is output. Through target calibration, the center point positions of the outer rectangles of each connected area are obtained. Then, by eliminating the target area, a binary image containing only the target area is obtained. If n represents the number of target areas, the target area is aggregated as $L = \{l_1, l_2, \dots, l_n\}$, and the distance threshold is set as σ . The distance between adjacent target areas is calculated using formula (11). If the distance between two adjacent target areas is less than or equal to σ , the two target areas will be merged and the center point coordinates of the merged target area will be recalculated. If it is greater than σ , adjacent target areas will be re selected to determine whether to merge until all adjacent target areas have a distance greater than σ . In this algorithm, the center position of the target area can be determined, so the motion speed of the same moving target in adjacent frame images can be calculated using formula (14).

$$v = \sqrt{(x_{1j_ctr} - x_{2j_ctr})^2 + (y_{1j_ctr} - y_{2j_ctr})^2} \quad (14)$$

In formula (14), (x_{1j_ctr}, y_{1j_ctr}) and (x_{2j_ctr}, y_{2j_ctr}) represent the center position of the j -th moving target in adjacent images [17-18]. The target tracking algorithm is not only related to the position and speed of the target, but also to changes in the environment. Therefore, a new target tracking algorithm has been studied and designed, and its process is shown in Fig. 5.

After inputting the initial parameter attribute set $object$, an image sequence of size $I \times J$ is output. The image frame counter is set to 0, various parameters of $object$ are initialized, and then the number of targets is reset. $k = l + 1$, the target region synthesis algorithm and formula (14) are used to calculate the positions and motion velocities of all targets in frame k . If the image frame counter is 0, the position and velocity of targets in frame $object$ are updated. If it is greater than 0, the position of targets in frame $object$ is used to predict the position of targets in frame k . Then, formula (15) is used to calculate the distance d between the target position and the predicted position.

$$d = \sqrt{(x_0 - x_m)^2 + (y_0 - y_m)^2} \quad (15)$$

In formula (15), x_0, y_0 represents the predicted position and x_m, y_m represents the calculated position. At this point, if d is less than the threshold, the target in $object$ matches the target in frame k of the image. The target position in $object$ is updated using the target position in the image until the image target leaves the visual area and all frames of the video image have been detected. This algorithm has improved the input parameter quality of the neural network, but has not improved the neural network and warning scheme. Therefore, the model is only suitable for a single moving target and cannot detect the motion direction of multiple targets. Thus, a new object tracking and collision warning algorithm has been studied and designed, and its process is shown in Fig. 6.

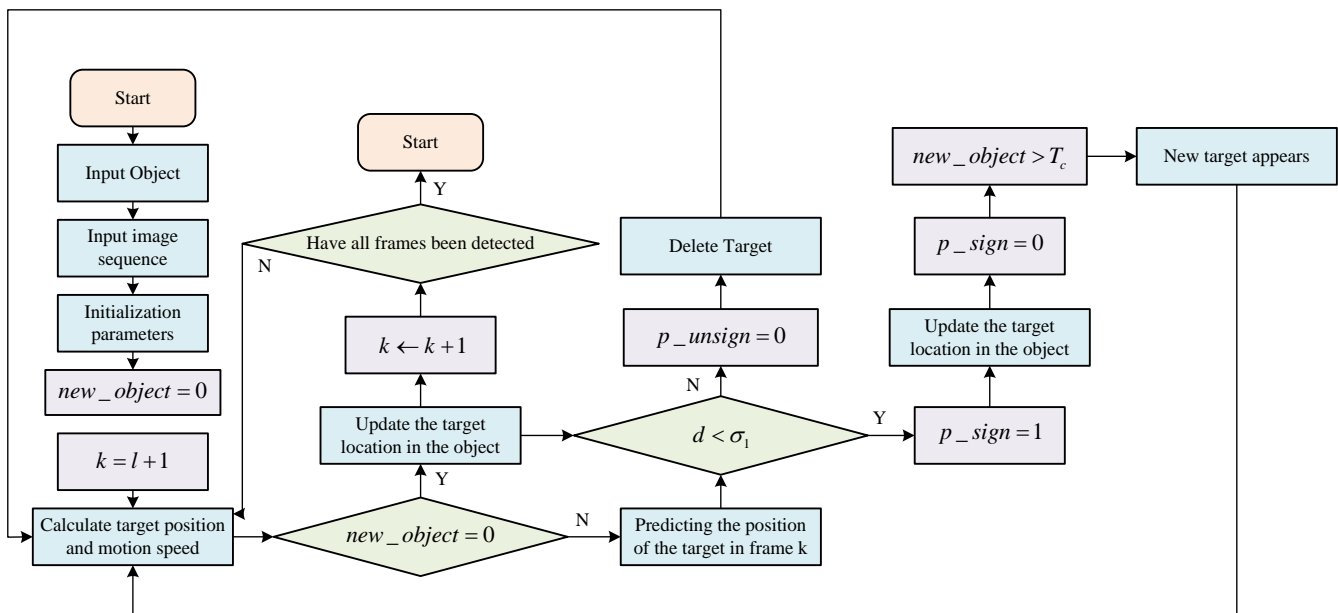


Fig. 5. Target tracking algorithm flow.

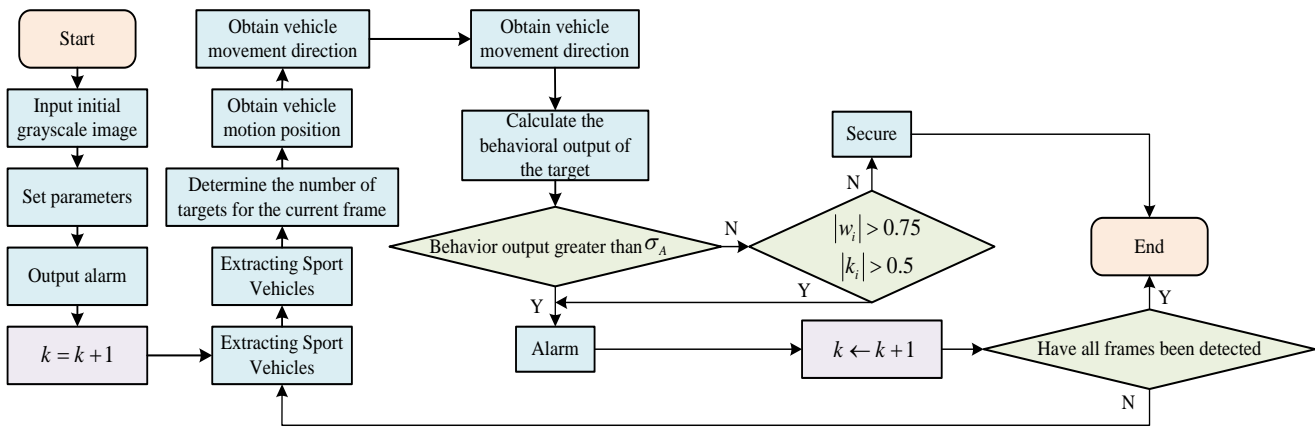


Fig. 6. Target tracking and collision warning algorithm flow.

An initial grayscale image of size $I \times J$ is input, the dynamic threshold σ_A and warning threshold frame number T of the algorithm are set, the warning signal is output. $k = l + 1$, the grayscale image of frame k is input, and the moving target of the image is extracted. The target area synthesis algorithm is used to determine the number of moving targets in the image, and then the target tracking algorithm is used to obtain the vehicle's motion position and direction. After obtaining the motion information of the target in the image, an artificial *Drosophila* visual neural network is used to calculate the behavior output of a vehicle, if the behavior output of the vehicle is greater than σ_A , a warning will be issued. If the behavior output of the vehicle is less than σ_A , the slope of the moving vehicle's motion direction will be calculated. If the absolute value of the moving vehicle's motion direction is greater than 0.75, and the absolute value of the slope is greater than 0.5, a warning signal will be sent. Finally, $k \leftarrow k + 1$, the behavior output of the vehicle can be obtained and compared with σ_A until all image frames are detected.

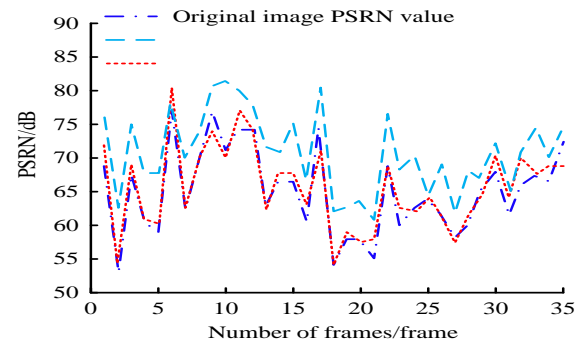
IV. SIMULATION EXPERIMENTAL RESULTS

This Section is an experimental validation analysis of the content of Section II, divided into two sections. The first section is the experimental validation analysis of improving the *Drosophila* visual neural network collision detection model, and the second section is the experimental validation analysis of target tracking and vehicle collision warning algorithms. The experimental validation analysis of improving the collision detection model of the *Drosophila* visual neural network can verify the recognition accuracy of the proposed algorithm in different situations, and the experimental validation analysis of target tracking and vehicle collision warning algorithms can verify whether the algorithm proposed by the acting team is effective in protecting drivers and personnel.

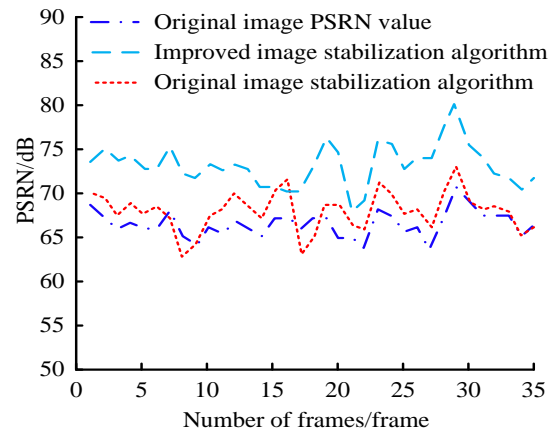
A. Collision Warning Experiment

To verify the feasibility of improving the *Drosophila* visual neural network collision model through research, experimental research was conducted under the OpenC V2.4.9 configuration in the Visual Studio 2010 software of the Windows 7 system. The study compared the Peak Signal to

Noise Ratio (PSNR) values of the image stabilization images obtained by the improved and original image stabilization algorithms, as shown in Fig. 7.



(a) PSNR value of video screen 1 and stable image



(b) PSNR value of video screen 2 and stable image

Fig. 7. Comparison of image stabilization algorithm results.

Fig. 7(a) shows the comparison of the PSNR values of the stabilized image of Video 1. The PSNR values of the improved stabilized image algorithm intersected with the PSNR values of the other two stabilized images, but overall they were still significantly better than the PSNR values of the other two stabilized images. The PSNR values of the improved stabilized image algorithm were up to 82dB and

down to 60dB, the PSNR values of the original image were up to 77dB and down to 52dB, and the PSNR values of the original stabilized image were up to 80dB and down to 54dB. Fig. 7(b) shows the comparison of the PSNR values of the stabilized image of Video 2. The improved stabilized image algorithm had significantly higher PSNR values than the other two stabilized images. In frame 16, the original stabilized image algorithm had a PSNR value of 71dB, which was higher than the improved stabilized image algorithm's 70dB. In the other frame numbers, the improved stabilized image algorithm had a PSNR value that was significantly higher than the original stabilized image, indicating a significant advantage. After confirming the effectiveness of the research in improving the image stabilization algorithm, the study compared the detection results of the Drosophila visual neural network collision detection algorithm with the improved Drosophila visual neural network algorithm, as shown in Fig. 8.

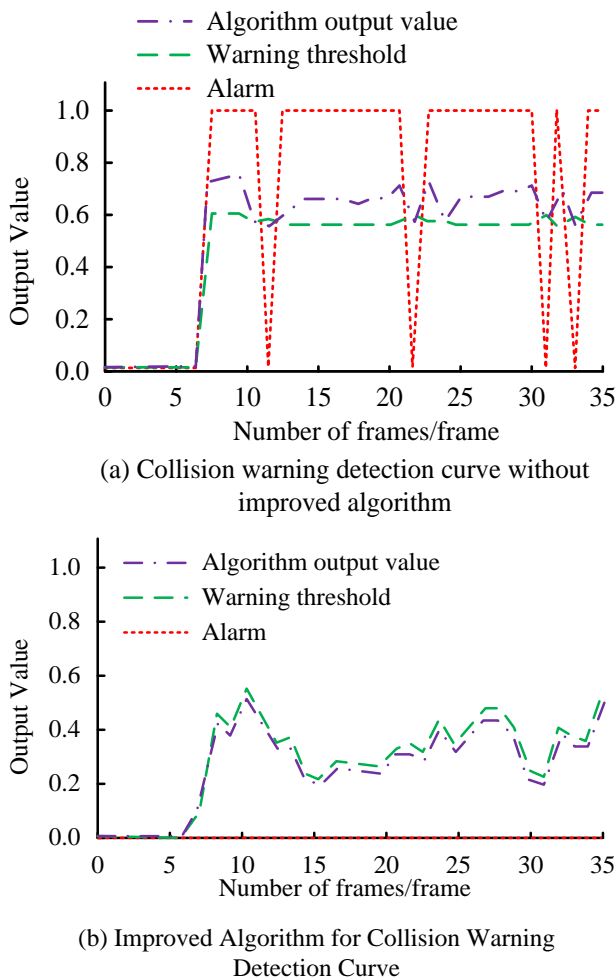


Fig. 8. Collision warning detection results of video screen 1.

Fig. 8(a) shows the collision warning detection results of the unimproved algorithm. Only at frames 12, 22, 31, and 34, the output value of the algorithm was less than the alarm threshold, and in the remaining frames, the output value of the algorithm was higher than the threshold, requiring an alarm.

Fig. 8(b) shows the collision detection probability results of the improved algorithm. The output value of the algorithm was always below the warning threshold, and there was no need to alarm. Based on the video content, a moving vehicle appears in frame 7 of the video. Afterwards, the moving vehicle began to move away from the camera, and a safe distance was maintained between the moving vehicle and the camera without warning. Therefore, the false alarm situation of the unimproved algorithm was relatively serious, while the improved algorithm did not show any false alarms. The detection results of the two algorithms in video 2 are shown in Fig. 9.

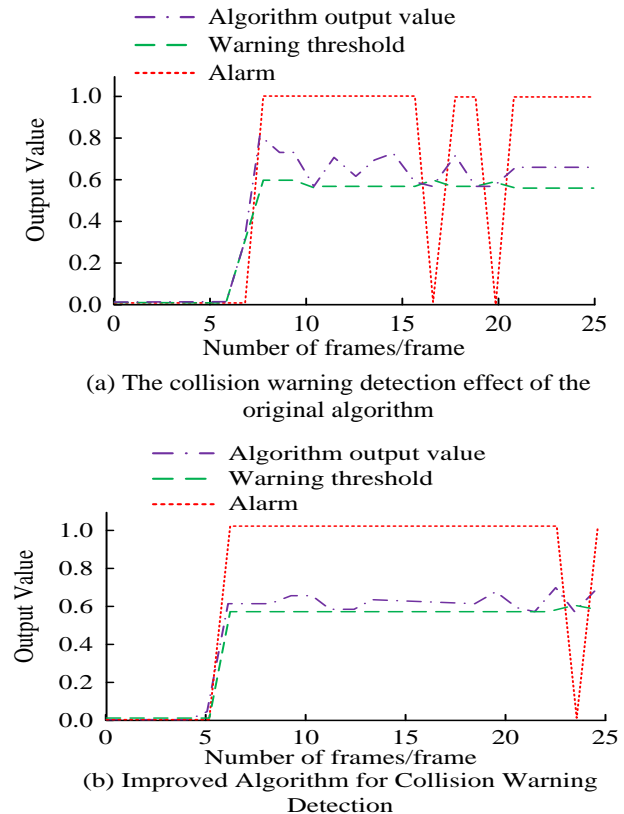
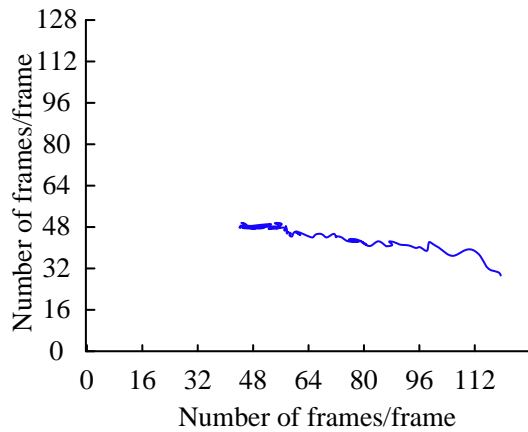


Fig. 9. Collision warning detection results of video screen 2.

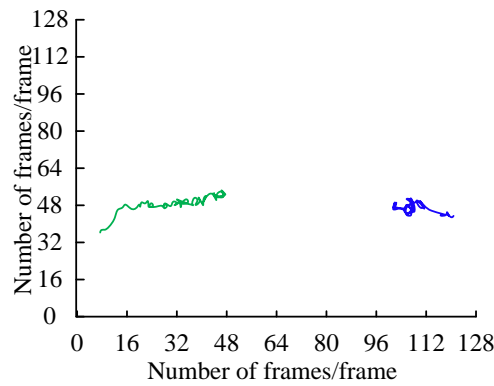
Fig. 9(a) shows the collision warning detection results of the unimproved algorithm. Starting from frame 6, the alarm threshold started to rise. At frame 8, when the algorithm output value exceeded the alarm threshold, the algorithm started to alarm. At frames 17 and 20, when the algorithm output value was lower than the alarm threshold, the alarm was canceled. Fig. 9(b) shows the collision warning detection results of the improved algorithm. Starting from frame 5, the alarm threshold rose to around 0.6 and stabilizes. At this point, the algorithm output value was higher than the alarm threshold, and the algorithm remains in an alarm state. Based on the content of video 2, the moving vehicles were gradually approaching the camera, and the predicted results of this settlement method were more in line with the actual results.

B. Collision Warning Effect

Before verifying the collision warning effect, the research first verified the feasibility of target calibration and tracking algorithms to obtain the motion curve of moving targets in the video sequence. The results are shown in Fig. 10.



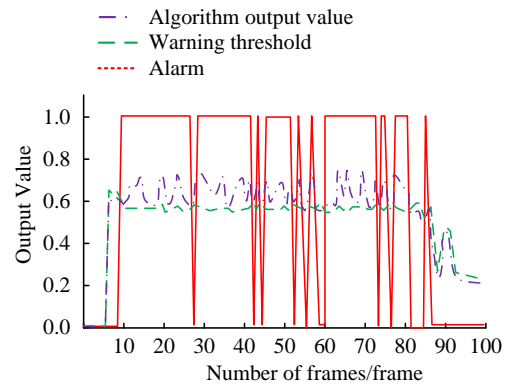
(a) Target motion trajectory of video screen 1



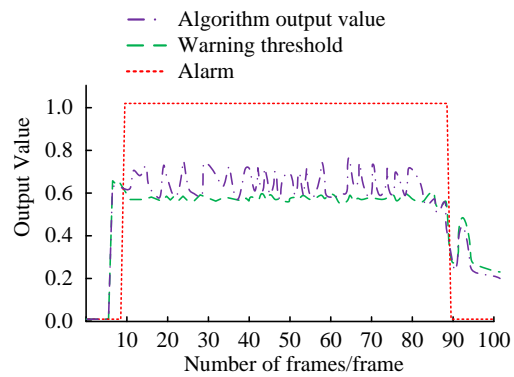
(b) Target motion trajectory of video screen 2

Fig. 10. Motion curves of vehicles with targets in two different types of videos.

Fig. 10(a) shows the tracking curve of the moving target in video 1. In video 1, there was only one target vehicle, and the tracking curve gradually moved from the right side towards the center of the camera's visual area, indicating that the target gradually moved away from the camera from the right side, which was consistent with the actual situation in the video. Fig. 10(b) shows the tracking curve of the moving target in video 2. There were two target vehicles in video 2. The motion curve of the first moving target gradually moved towards the center of the camera's visual area from the left, while the second vehicle hovered on the right side of the camera's visual area, which was consistent with the actual situation in the video. After extracting the target motion curve, the study compared the traditional collision warning model with the collision warning model designed in the study. The warning effect in two video scenarios is shown in Fig. 11 for video 1.



(a) The collision warning detection effect of the original algorithm



(b) Improved Algorithm for Collision Warning Detection

Fig. 11. Comparison of two algorithms for warning effects in video 1.

Fig. 11(a) shows the warning effect of the traditional collision warning algorithm. At frame 6, the moving target began to appear, and the algorithm started to alarm until frame 89. During frames 9 to 89, there were multiple occurrences of non-alarm areas, and the algorithm output values were all above the alarm threshold. Fig. 11(b) shows the collision warning effect of the research and design algorithm. From frame 9 when the moving target appeared to frame 89, the moving target left the visual area, the algorithm was in an alarm state without any missed alarms. The warning effect of Video 2 is shown in Fig. 12.

Fig. 12(a) shows the warning effect of traditional collision warning algorithms. During frames 1 to 18, the moving target appeared, and the algorithm output value and alarm threshold fluctuated. Starting from frame 22, the algorithm started to intermittently alarm, and the overall output value of the algorithm remained above the warning threshold. Fig. 12 (b) shows the collision warning effect of the research and design algorithm. The changes in the algorithm output value and warning threshold from frames 1 to 18 were consistent with traditional algorithms. Since frame 22, the research and design algorithm was always in a warning state. Based on the analysis of video content, traditional algorithms suffered from serious false positives, while research and design algorithms showed no false positives.

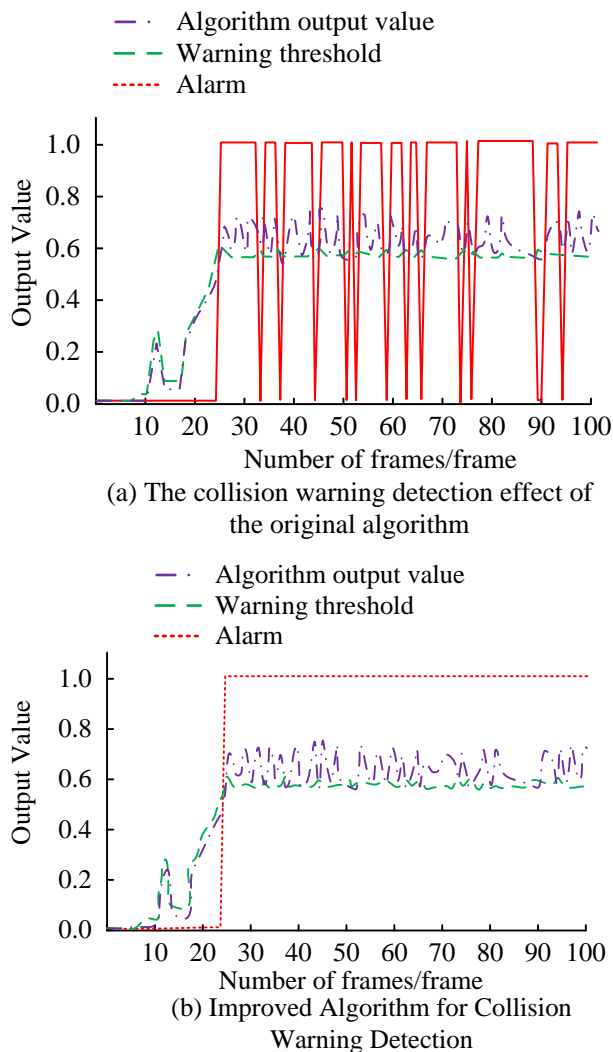


Fig. 12. Comparison of two algorithms for warning effects in video 2.

V. CONCLUSION

To increase the safety of vehicle driving and ensure the safety of drivers' lives and property, a vehicle tracking and collision warning model based on improved Drosophila vision is studied, designed, and constructed. The model adopts motion vector estimation algorithms, image stabilization algorithms, target region synthesis algorithms, target tracking algorithms, and collision warning algorithms that are improved based on traditional algorithms. The results showed that the PSNR value of the improved image stabilization algorithm was always higher than that of the original image and traditional image stabilization algorithms. The PSNR value of the improved image stabilization algorithm was highest at 82dB and lowest at 60dB, while the PSNR value of the traditional image stabilization algorithm was highest at 80dB and lowest at 54dB. The improved Drosophila visual neural network collision warning algorithm had no false positives, and the original algorithm had four false positives in video 1. The research and design of the target region synthesis algorithm can complete the task of extracting motion curves of video moving targets. The algorithm extracted the motion

curves of one target in video 1 and two targets in video 2, which were consistent with the video content. The target tracking and collision warning algorithm designed in the study did not show any false positives. In the collision warning effects of Video 1 and Video 2, the original algorithm showed relatively serious false positives, while the algorithm designed in the study did not, which was consistent with the actual information in the video. The algorithm designed through research has improved the missed and false positives of the original algorithm, but the improved image stabilization algorithm has poor performance when facing multiple objectives. Subsequent research can continue to optimize the application effect of the model in multi-objective situations.

REFERENCES

- [1] C. Pek, S. Manzinger, M. Koschi, and M. Althoff, "Using online verification to prevent autonomous vehicles from causing accidents," *Nat. Mach. Intell.*, vol. 2, no. 9, pp. 518–528, Sept. 2020.
- [2] A. Borucka, E. Kozowski, P. Oleszczuk, and A. Widerski, "Predictive analysis of the impact of the time of day on road accidents in Poland," *Open Eng.*, vol. 11, no. 1, pp. 142–150, Dec. 2020.
- [3] W. P. Sanberg, G. Dubbelman, and P. With, "ASTEROIDS: A stixel tracking extrapolation-based relevant obstacle impact detection system," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 4, pp. 34–46, Mar. 2020.
- [4] J. Hu, Y. Zhang, and S. Rakheja, "Path planning and tracking for autonomous vehicle collision avoidance with consideration of tire-road friction coefficient," *IFAC-PapersOnline*, vol. 53, no. 2, pp. 15524–15529, Jul. 2020.
- [5] F. Cao, and H. Jiang, "Trajectory planning and tracking control of unmanned ground vehicle leading by motion virtual leader on expressway," *IET Intell. Transp. Syst.*, vol. 15, no. 2, pp. 187–199, Dec. 2020.
- [6] X. Zhou, X. Yu, Y. Zhang, Y. Luo, and X. Peng, "Trajectory planning and tracking strategy applied to an unmanned ground vehicle in the presence of obstacles," *IEEE Trans. Autom. Sci. Eng.*, vol. 18, no. 4, pp. 1575–1589, Aug. 2020.
- [7] F. Tokuda, S. Arai, and K. Kosuge, "Convolutional neural network-based visual servoing for eye-to-hand manipulator," *IEEE Access*, vol. 9, no. 6, pp. 91820–91835, Jun. 2021.
- [8] A. Burguera and F. Bonin-Font, "An unsupervised neural network for loop detection in underwater visual SLAM," *J. Intell. Robot. Syst.*, vol. 100, no. 3/4, pp. 1157–1177, Aug. 2020.
- [9] Q. Gu, J. Su, and L. Yuan, "Visual affordance detection using an efficient attention convolutional neural network," *Neurocomputing*, vol. 440, no. 14, pp. 36–44, Jun. 2021.
- [10] Q. Fu, C. Hu, J. Peng, F. C. Rind, and S. Yue, "A robust collision perception visual neural network with specific selectivity to darker objects," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 5074–5088, Dec. 2020.
- [11] J. Tao, C. Yang, and C. Xu, "Estimation algorithm of incident sources' stokes parameters and 2% DOAs based on reduced mutual coupling vector sensor," *Radio Ence*, vol. 54, no. 7/8, pp. 770–784, Aug. 2019.
- [12] W. Choi, I. Y. Song, and V. Shin, "Two-stage algorithm for estimation of nonlinear functions of state vector in linear Gaussian continuous dynamical systems," *J. Comput. Syst. Sci. Int.*, vol. 58, no. 6, pp. 869–882, Feb. 2019.
- [13] M. D. Hua, J. Trunpf, T. Hamel, R. Mahony, and P. Morin, "Feature-based recursive observer design for homography estimation and its application to image stabilization: Feature-based Recursive Observer Design for Homography Estimation," *Asian J. Control*, vol. 21, no. 4, pp. 1443–1458, Jan. 2019.
- [14] T. Rasal, T. Veerakumar, B. N. Subudhi, and S. Esakkirajan, "Mixed poisson gaussian noise reduction in fluorescence microscopy images using modified structure of wavelet transform," *Image Process., IET*, vol. 15, no. 7, pp. 1383–1398, Dec. 2020.

- [15] Z. Chen, "Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm," *J. Comput. Cogn. Eng.*, vol. 1, no. 3, pp. 103-108, Mar. 2022.
- [16] A. N. Sharkawy, P. N. Koustoumpardis, and N. Aspragathos, "Neural network design for manipulator collision detection based only on the joint position sensors," *Robotica*, vol. 38, no. 10, pp. 1737-1755, Jun. 2020.
- [17] X. Yang, S. Zhu, D. Zhou, and Y. Zhang, "An improved target tracking algorithm based on spatio-temporal context under occlusions," *Multidimension. Syst. Signal Process.*, vol. 31, no. 1, pp. 329-344, Jun. 2020.
- [18] X. Fang and L. Chen, "Noise-aware manoeuvring target tracking algorithm in wireless sensor networks by a novel adaptive cubature Kalman filter," *IET Radar, Sonar Navig.*, vol. 14, no. 11, pp. 1795-1802, Sept. 2020.

Earth Observation Satellite: Big Data Retrieval Method with Fuzzy Expression of Geophysical Parameters and Spatial Features

Kohei Arai
Information Science Dept.
Saga University
Saga City, Japan

Abstract—A method for fuzzy retrievals of Earth observation satellite image database using geophysical parameters and spatial features is proposed. It is confirmed that the proposed method allows fuzzy expressions of queries with sea surface temperature, chlorophyll-a concentration and cloud coverage as well as circle, line and edge, for instance “rather cold sea surface temperature and a sort of circle feature”. Thus users, in particular, oceanographers may access the most appropriate image data from the database for finding of cold cores (circle features), fronts (arc and line features), etc. in a simple manner. Although this is just an example for oceanographers, it is found that the proposed method allows data mining with fuzzy expressions of geophysical queries from the big data platforms of the earth observation satellite database.

Keywords—Fuzzy retrieval; earth observation satellite; big data; geophysical parameter; oceanographer; circle feature; arc feature; line feature; fuzzy expression

I. INTRODUCTION

These remote sensing satellite data are totally big data. One of the problems on the big data analysis is how to retrieve most appropriate satellite data. In this paper, Earth observation satellite data retrieval method with fuzzy expression of geophysical parameters and spatial features is proposed.

Conventional search engine allows search objects with some multimedia of keywords, images, voices, etc. In terms of Earth observation data retrieval, there is strong demand on satellite imagery data search with geophysical features and spatial features (for instance, “rather cold sea surface temperature and a sort of circle feature”). Such this flexible data search engine is required for the remote sensing imagery data users.

Space development agencies (NASA, NOAA, NASDA, ESA, IRS, CNES, etc.) that are building databases of earth observation satellite data have developed WWW or proprietary search software as a user interface for database search and provide it to users. Search requests from these are translated to be effective in databases based on unique structures and languages (most institutions are Oracle databases (relational databases) [1] and are translated into search requests in SQL language) [1].

The database consists of the earth observation satellite image data to be searched (the unit of search is called a

granule), the inventory data that is its catalog information, and the browse image data with reduced resolution to know the outline of the database (this is called browsing) and their location information and metadata as granule attribute information.

To search for a granule, first, the inventory is searched from the inventory database, and the browse image data that matches the search conditions is searched from the browse image database via the metadata in the meta database. From the database, since location information and the like are linked between these databases, the user only needs to set conditions for retrieving inventory and browse images. In the current space development agency database system, the conditions for inventory search start with the satellite name, sensor name, ground station name, and indicate the observation date and time, location, data quality, cloudiness, etc.

From the inventory data candidates that meet these conditions, the one closest to the desired one is selected, the browse image is displayed and confirmed, and this is repeated until the desired data is reached. However, since neither the inventory data nor the browse image contains information on the physical quantity, which is the evaluation criterion for the user data, the number of repetitions until the desired data is reached is not small.

The retrieval method of the Earth Observation Satellite Image Database: EOSID proposed in this paper aims to reduce the time required for retrieval by adding physical quantities and spatial features of images to inventory data. There are many physical quantities and spatial features of earth observation satellite image data, but here I limited them to those in the marine field as an example. In other words, sea surface temperature and chlorophyll-a concentration are used as physical quantities, and edges, lines, circles, and arcs are selected as spatial features, and based on these quantities, The search conditions were narrowed down.

Furthermore, when specifying these physical quantities and spatial features as search conditions, it is assumed that it is impossible to give them in a limited manner, and a search based on fuzzy theory with a very, slightly, etc., language hedge should be used. This search based on fuzzy theory with language hedging has already been proposed by, for example, Isomoto [2], and is not new at all. And the search method must

be devised. For example, Sobue et al. have proposed a method for searching catalog information of earth observation satellite data using physical quantities as fuzzy search targets [3].

NASA has also applied this to text search for data and information on the global environment. Here, the author proposes a method of further narrowing down the search items by referring to the spatial features by further developing [3]. The author reports the effectiveness of the proposed search method using a virtual Earth observation satellite image database.

In the following section, research background and related research works are described. Then, the proposed method and system is described followed by experimental set-up together with experimental results. After that, concluding remarks and some discussions are described.

II. RESEARCH BACKGROUND

The demand for satellites related to remote sensing has grown significantly, especially in emerging countries where many countries do not have launch vehicles. In emerging countries, there is about four times the demand in the last ten years and the next ten years. It is announced that Metaps, who supports app monetization using artificial intelligence, will start joint research on a big data analysis system using micro satellites in cooperation with space shift. Many new and old players (Google, Facebook, etc.) are promoting market and customer development with various approaches such as resolution, shooting frequency, analysis, cost, etc., the global satellite remote sensing market.

Microsoft and the United States Ocean Atmosphere Agency (NOAA), a joint R & D agreement to develop the best way to extract data from internal systems. This will allow Microsoft to provide weather, water, and oceans provided by NOAA scientists and weather data hosted on Azure cloud platform. This collaboration is an important step in realizing the promise of open data and innovation, which allows governments and businesses to leverage NOAA data aggregated in cloud repositories and then availability of large amounts of computing resources Partners and customers develop new solutions for citizens and customers. Digital Globe, a satellite image provider (34 cm spatial resolution) founded in 1992, signed a long-term contract with the US government in 2002, and merged with GeoEye in 2012, with sales of approximately 60 billion JYen. The company is a major customer of NGA (National Geospatial-Intelligence Agency), and the government business accounts for over 80% of sales. On the other hand, vendors providing mapping services such as Google and Nokia are also customers. Skybox Imaging, which was acquired by Google, will consider releasing satellite data, combining satellite HD video data with map APIs to instantly see the movement of the ground "satellite video". It can be used to monitor the movement of airplanes and ships, and to identify illegal deforestation, etc.

Planet Labs is a company that sells satellite photos collected from a network of 87 small satellites, and with this acquisition, RapidEyes has a six-year archive of six billion square kilometers of global land images, More than 177 Dove satellites are launched on orbit, and a large number of

constellations that shoot the same spot at least once a day (multiple earth observation satellites are thrown in the same orbit, high frequency observation and then provided free of charge by Creative Commons (CC) license Start. The CC license is a new copyright rule for the Internet age, a tool for the author of publishing works to indicate that "you are free to use my work if you comply with these conditions."

Spire, a satellite venture, uses a group of small satellites called satellite constellations to collect various data and analyze geospatial information such as world trade and weather.

Descartes Labs mainly does the analysis of satellite images to understand what is there and to educate the system to extract specific crops from satellite images, extracting significant data from them, Adapt to the yearly satellite imagery. Predict actual yield by applying statistical model to extracted crops. Specifically, take out corn grown on the farm from satellite image and predict the amount of harvest for that year.

Facebook embarks on accurate mapping of the world by using artificially captured images (remote sensing images) and artificial intelligence technology (AI). How many people live in which region of which country know exactly what it is and optimize its global broadband offering, the Connectivity Lab uses AI technology for approximately 14.6 billion satellite images in 20 countries around the world to identify man-made structures etc., how many houses are along the river and along the road, and what communities We analyzed what was formed. Efforts to use "Computer Vision: Business Intelligence" where CIA-affiliated companies can work with Amazon to peek into Earth information at an unprecedented level of detail.

Omni Earth is planning a satellite system consisting of 18 units, and in the future, it will assume an earth observation data volume of 60 petabytes /year. The company is characterized by a partnership, and in 2014, partnered with Dynetics for satellite design and manufacturing, Draper Laboratory for systems engineering, and Spaceflight for launch service. The company emphasizes solution development using satellite image analysis, and acquired the US IRIS maps in August this year, and launched a business solution integrating earth observation images and other data for agriculture, forestry, energy and public sector provide on a cloud basis. The combination of advanced satellite infrastructure and advanced application development can create new innovations.

Services such as Google Maps, Microsoft's Bing, and MapQuest will display various satellite images cut and pasted. On the other hand, in the map box, we use the vast array of satellite image data of NASA and take an approach of stacking innumerable images taken from one area as a layer, to realize a clear image without a seam. A similar framework was born in Europe in February 2016. ESA concludes LOI (Letter of Intent) with SAP for rapid and efficient utilization of huge earth observation data. As satellite observation data by the Earth observation program "Copernicus" advanced by ESA is huge, data processing is difficult with conventional technology. Build an innovative approach to data processing and analysis by leveraging SAP's cloud platform "SAP HANA Cloud". ESA launches TEP and promotes practical use in six fields, Disaster

prevention, Coast, forest, Water resources, Polar region, Cities and infrastructure. One example (coastal): Supporting the Aquaculture Fisheries Industries (SAFI) is a fishery data server for aquaculture companies and fishermen. Such these platforms provide kinds of big data. More importantly, data mining from the big data platform has to be done in efficient and effective manner. The proposed method would like to provide a sophisticated manner of data mining with fuzzy expression of geophysical parameters.

III. RELATED RESEARCH WORKS

Vague search of earth observation image database based on Fuzzy theory using physical quantities and spatial features is proposed [4] together with earth observation satellite image database system allowing ambiguous search requests [5]. On the other hand, user friendly and efficient catalog information management for earth observation data is proposed and well reported [6].

Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data is proposed [7]. Meanwhile, data collection and active database for tsunami warning system is proposed [8].

A review of Chinese Academy of Science (CASIA) gait database as a human gait recognition dataset is conducted [9] together with gait recognition method based on wavelet transformation and its evaluation with CASIA gait database as human gait recognition dataset [10]. Meanwhile, visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals is proposed and validated [11] together with visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals [12].

Wavelet based image retrieval method is proposed and evaluated its usefulness [13]. On the other hand, DP matching based image retrieval method with wavelet Multi Resolution Analysis: MRA which is robust against magnification of image size is proposed [14]. Meanwhile, Free Open-Source Software: FOSS based Geographic Information System: GIS for spatial retrievals of appropriate locations for ocean energy utilizing electric power generation plants is proposed [15].

Error analysis of air temperature profile retrievals with microwave sounder data based on minimization of covariance matrix of estimation error is conducted [16]. Meanwhile, visualization of link structure and URL retrievals utilization of interval structure of URLs based on brunch and bound algorithms is well reported [17]. Method for image portion retrieval and display for comparatively large scale of imagery data onto relatively small size of screen which is suitable to block coding of image data compression is also proposed [18].

Content based image retrieval by using multi-layer centroid contour distance is proposed [19]. On the other hand, remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data is proposed [20].

Image retrieval and classification method based on Euclidian distance between normalized features including

wavelet descriptor is proposed [21]. Also, numerical representation of web sites of remote sensing satellite data providers and its application to knowledge-based information retrievals with natural language is proposed [22]. Image retrieval based on color, shape and texture for ornamental leaf with medicinal functionality, meanwhile, is proposed [23]. Also, comparison contour extraction based on layered structure and Fourier descriptor on image retrieval is proposed and evaluated its effectiveness [24].

Pursuit Reinforcement Competitive Learning: PRCL-based online clustering with tracking algorithm and its application to image retrieval is proposed [25]. Also, image retrieval method utilizing texture information derived from Discrete Wavelet Transformation: DWT together with color information is proposed [26].

Metadata definition and retrieval of earth observation satellite data is proposed [27]. On the other hand, Open GIS with spatial and temporal retrievals as well as assimilation functionality is proposed [28]. Meanwhile, Geographic Information System: GIS based on neural network for appropriate parameter estimation of geophysical retrieval equations with satellite remote sensing data is proposed [29].

Image retrieval method based on hue information and wavelet description-based shape information as well as texture information of the objects extracted with dyadic wavelet transformation is proposed [30]. Wavelet based image retrievals is attempted [31]. Also, image retrieval method based on back projection is proposed [32].

For the Fuzzy logic related research works, there are the following papers,

Fuzzy Genetic Algorithm (GA) for prioritization determination with techniques for order performance by similarity to ideal solution is proposed [33]. On the other hand, smart grid photovoltaic system pilot scale using sunlight intensity and state of charge (SoC) battery based on Mamdani fuzzy logic control is also proposed [34]. Satellite image database search engine which allows fuzzy expression of geophysical parameters of queries is proposed [35]. Meanwhile, operation of light tracker movement using Fuzzy logic control information and communication technology is proposed [36].

IV. PROPOSED METHOD

A. Search Procedure

The search procedure in the current database system of most space development agencies is as follows:

- 1) As a condition for inventory search, start with satellite name, sensor name, ground station name, and specify observation date and time, location, data quality, cloudiness, etc. The user inputs a standardized search key according to a request from the terminal.
- 2) Search for inventory information based on the input search conditions.
- 3) Search for metadata corresponding to the searched inventory information.

4) Based on the retrieved metadata, a browse image that matches the search condition is retrieved from the database where it is located and presented to the user.

5) If there is no browse image desired by the user, the process returns to step 1. If there is a desired browse image, the original image data corresponding to the selected browse image is further searched.

6) Return the searched image data to the user as a search result.

In contrast, the search procedure proposed here is as follows:

1) *Inventory search conditions include:* satellite name, sensor name, ground station name, observation date and time, location, data quality, cloudiness, sea surface temperature, chlorophyll-a concentration, and other physical quantities and edges, lines, circles, and arcs. And other spatial features. At this time, for the search condition items below the cloud cover, search requests with ambiguous expressions by language hedging are allowed.

2) Search for inventory information based on the input search conditions.

3) Analyze the given ambiguous search condition and find the threshold that matches the condition using fuzzy logic.

4) If there is an image in the database that contains data between the threshold value calculated in 3 and the maximum value of the target membership function, it is assumed that the condition is met and the result is retained. .

5) Process image data with data that meets all the given conditions to create a browse image and display it as a candidate for search results along with attribute data such as sea surface temperature of the image.

6) The original of the browse image data selected by the user display the image data as a result. That is, the points different from the current search method are listed as follows.

a) The physical quantities and spatial features are included in the inventory data, and search conditions can be set based on these.

b) The use of attribute information included in the browse image database to confirm the matching of search conditions for physical quantities and spatial features eliminates the need for a meta database.

c) It is allowed to set ambiguous search conditions for these physical quantities and spatial features.

B. Setting Search Key Items

Since the earth observation field is wide-ranging, the ocean observation data is taken up with particular attention to the ocean field. The concept of the basic search method is common to each field. The search key items used this time are as follows,

1) Sea surface temperature, chlorophyll-a content, cloudiness,

2) Number of edges, number of circles, number of arcs, number of lines from spatial features.

The author chose, in particular, by enabling search using this spatial feature, it is thought that it is useful for search of the earth observation satellite image database by ocean dynamics researchers, topographical geological researchers, etc. For example, those who study ocean dynamics pay attention to the tides and the shape of the current axis when watching the movement of warm and cold currents. This tide is located in the convergence zone of the ocean current, where the temperature difference of seawater is sharp.

The sea conditions in this tidal sea area fluctuate greatly both temporally and spatially, and both hot and cold-water masses are disturbed and local convergence, subsidence areas, divergence, and upwelling areas are complicatedly arranged. In this area, the fishery is generally rich in nutrients and high in productivity and it is easy to gather both cold and warm schools of fish due to the flow, making it a good fishing ground. When this tide is captured as data, there are various shapes such as line, arc, circle, line pair, and the like. For example, looking at a line, there are various factors such as size, angle, and location. There are several of these shapes in one tide. To look at the tide from various angles and know the movements of the warm and cold-water flows, it is necessary to extract the necessary tide locations and images from the database.

In addition, if it becomes possible to search using the spatial characteristics of circles, it can be expected that cold water chunks, etc., which are known for forming good fishing grounds, can be easily searched. As other search key items, the observation place and observation date adopted by the current

Earth observation satellite image database systems were used. These search key items and various quantities of the browse image to be searched are as follows,

- 1) Image number << No. 1 ~ No. 100 >>
- 2) Sea surface temperature << 0-30 degrees >>
- 3) Chlorophyll-a content (0 mg / m³-35mg / m³)
- 4) Number of edges << 0 ~ 14 >>
- 5) Number of circles << 0 ~ 4 >>
- 6) Number of arcs << 0 ~ 5 >>
- 7) Number of lines << 0 ~ 4 >>
- 8) Observation place << random >>
- 9) Observation date << random >>
- 10) Cloud Cover << 1% ~ 100% >>

As described above, one browse image holds ten attribute values.

C. Search Engine Based on Fuzzy Theory

1) *Membership function:* In order to perform a search based on ambiguous search requests from users, a membership function is defined using fuzzy theory that can handle ambiguity. Assemble the search conditions using the defined membership function and perform the search together with observation location, observation date, etc.

It is considered that the analyst to search for is determined by the amount of cloud and is not specified vaguely. Therefore, the author considers a search method that allows ambiguous expressions for the remaining seven items. Therefore, it is

necessary to define a membership function for each physical quantity and space feature that we have taken up this time. However, how people perceive certain terms is different. Therefore, there is a problem that the membership function cannot be determined uniquely. Therefore, here, a questionnaire survey was conducted for members of the Information System Subcommittee of the Forum of the Earth Science and Technology Agency, and the membership function of each search condition item was determined. The exponential function of Eq. (1) was used as the function system.

$$\mu(x) = \int_a^b e^{-c(x-d)^2/x} dx \quad (1)$$

where, x is the value of each search condition item, and a, b, c, d are coefficients of the membership function shown in Table I.

2) *Definition of the language hedge*: The language hedge is a modifier attached to attribute information and is an operator that converts into attribute information (fuzzy set) having a qualifying meaning. The language hedge used this time is as follows:

- a) Pretty
- b) Very (Very)
- c) To some extent (A sort of)
- d) Somewhat (Rather)
- e) Some (More or less)
- f) Slightly

TABLE I. COEFFICIENTS OF THE MEMBERSHIP FUNCTION FOR EACH KEYWORD

Search Keywords	Attributes	a	b	c	d
Sea Surface Temperature	Warm	0	30	0.015	30
	Cool	0	30	0.15	0
Chlorophyll-a Concentration	Concentrated	0	1	0.003	35
	Sparse	0	1	30.0	0
Cloud Coverage	Concentrated	0	100	0.034	17
	Sparse	0	100	3.0	0
No. of Edges	Many	0	14	0.5	5
	Little	0	14	0.32	0
No. of Circles	Many	0	5	0.25	5
	Little	0	5	0.15	0
No. of Arcs	Many	0	6	0.3	6
	Little	0	6	0.2	0
No. of Lines	Many	0	4	0.6	4
	Little	0	4	0.5	0

Membership function: Number of lines "less"

There are six types. In order to calculate the centroid value from the membership function during the search, these were defined as follows respectively,

- a) Pretty A = NORM (INT (A) and not TNT (CON (A)))
- b) Very A = CON (A)
- c) A sort of A = NORM (not CON (A) 2 and DIL (A))
- d) Rather A= NORM (TNT (A))
- e) More or less A = DIL (A)
- f) Slightly A= NORM (A and not (very A))

where, “A” means a fuzzy set, and not means a negation operation and a product operation. Also, CON, DIL, NORM, and INT are called centralization, enlargement, normalization, and contrast enhancement, respectively.

$$CON(A) = A^2, \mu_{con(A)}(x) = \mu^2(x), DIL(A) = \sqrt{A}, \quad (2)$$

$$\mu_{DIL(A)}(x) = \sqrt{\mu(x)}, NORM(A) = \frac{A}{\mu_{max}}, \quad (3)$$

$$\mu_{NORM(A)}(x) = \frac{\mu(x)}{\mu_{max}}, \quad (4)$$

$$\mu_{INT(A)}(x) = 2\mu^2(x), 0 < \mu(x) < 0.5 \\ = 1 - 2(1 - \mu(x))^2, 0.5 < \mu(x) < 1.0 \quad (5)$$

where, x is a search condition item (physical quantity and spatial feature).

3) *AND, OR, NOT in search condition*: In order to allow a combination of multiple different search conditions, a logical operation such as AND, OR, and NOT of search conditions was devised. Center of gravity value. When a search request is issued, the physical quantity and spatial shape data of each image are examined using a membership function corresponding to ambiguous conditions. However, since all data are examined, all data are search results. Therefore, a threshold is set, and if x is included between the threshold and the maximum value of the membership function, it is considered that the condition is met. This time, the threshold value adopted the value using the center of gravity (CG) of the fuzzy set. The center of gravity of the fuzzy set is calculated by the following equation,

$$CG = \int_x x \mu_A(x) dx / \int_x \mu_A(x) dx \quad (6)$$

If there is no data that satisfies the condition input by the user, the data closest to the barycentric value is output as a search result.

4) *Data and browse image search*: In advance, edges, circles, arcs, and lines, which are spatial features, were extracted from the earth observation satellite image, and an index was constructed. For the extraction of the spatial features of these images, we used a combination of simple image classification, edge extraction by the relaxation method, and thinning as a method to extract such spatial features.

For the form on the browser that actually performs input and output, select the form from which the search conditions are actually written from the keyboard, and the conditions that apply to the image you are looking for from the given word group I have prepared two formats that can be searched by pressing the button. In the former, the conditions are entered on the keyboard, and complicated conditions containing AND, OR, and NOT conditions can be entered. Therefore, the operation became a little difficult. On the other hand, the latter makes it possible for anyone to easily search for images and prepares the one that saves the trouble of inputting conditions as much as possible. Therefore, only one mode was selected from the AND, OR, and NOT conditions, and the entered condition was searched according to the mode.

From the homepage (Home Page of the proposed Earth observation satellite image database retrieval system), from which you can select either "text input" or "button input". If "text input" is selected here, it will transition to the input form (the menu for retrieval keyword input in text form) as shown in Fig. 1. Here, the conditions are assembled from the example shown, input from the keyboard, and the search is performed. When "button input" is selected, the screen shifts to the input form (Menu for retrieval keyword input by clicking mouse button) as shown in Fig. 2. In this case, the input method is simpler than "text input", but it is not possible to input complicated conditions with complicated conditions.

Text Input System
Available Items

A:	SST	1:	pretty	a:	Large
B:	Chlorophyll-a	2:	more_or_less	_	
C:	No. of_Edge	3:	very	_	
D:	No. of_Circle	4:	rather	b:	Small
E:	No. of_Arc	5:	normal	_	
F:	No. of_Line	6:	sort_of	_	
_	_	7:	slightly	_	
Operator:	and_or_not				

Fig. 1. Menu for retrieval keyword input in text form.

Please key in your keyword below,
Search

Example of combination SST is pretty high: A1a

Large chlorophyll-a concentration and (more or less of the number of edge or sort of the number of circle): B5a and (C2b or D4b)

V. EXPERIMENT

In this experiment, the browse image size was set to 64 by 64 pixels, and 100 images were prepared. These are a part of the original NOAA-11 / AVHRR¹ image of 3:37 GMT on April 25, 1990. The spatial features are extracted and compiled into a database by the method of reference, and the pixel interval of 1 km is also calculated. An 8 by 8 pixels average filter was used to create a browse image with a pixel interval of 8 km. Here, the condition "A7a and C3a (sea surface temperature is slightly warm, and the number of edges is very large)" is given as a search example.

Fig. 3 shows the example. An example of the retrieval for Sea Surface Temperature is slightly high and the number of edges is very large. The user can select a desired image from a given browse image and data of the image. An example of the retrieved results (Extract image No. 12 from the candidates of the browse images) shows the result of selecting image number 12. Fig. 4 shows an example of the retrieved results (Extract image No. 12 from the candidates of the browse images).

Sea Surface Temp. "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Sea Surface Temp. "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Chlorophyll-a "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Number of edge "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Number of circle "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Number of arc "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Number of line "Use" "Not use"

"Pretty" "more or less" "very" "rather" "normal" "sort of" "slightly"

"Large" "Small"

"AND" "OR"

Fig. 2. Menu for retrieval keyword input by clicking mouse button.

Searching A7a and C3a

=>Search Result

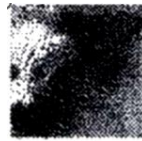
Image_No.	Browse	SST	Chlorophyll-a	Edge	Circle	Arc	Line
5		22	33.405	11	2	5	3
12		24	33.757	12	3	5	3
21		21	33.445	14	4	5	4
44		27	5.157	11	3	3	4
49		20	29.005	14	4	5	4
58		21	9.669	11	4	3	3
73		23	9.317	11	3	5	2

Submit Reset

Fig. 3. An example of the retrieval for sea surface temperature is slightly high and the number of edges is very large.

In addition, in the present search system, when there is no image satisfying the condition presented by the user, the image closest to the condition presented by the user is searched by gradually increasing the search width.

¹ https://en.wikipedia.org/wiki/Advanced_very-high-resolution_radiometer



The element of the image is as follows,

Image No.: 12
SST: 24
Chlorophyll-a: 33.757
No. of Edge: 12
No. of Circle: 3
No. of Arc: 5
No. of Line: 3

Fig. 4. An example of the retrieved results (Extract image No. 12 from the candidates of the browse images)

Fig. 5 and 6 show examples of browse images and their spatial features extracted as search results. Fig. 6 is obtained from the extracted lines, edges, arcs and circles from the image of Fig. 5 based on the relaxation method² with the thinning algorithm.



Fig. 5. Details of the retrieved browse image (64 by 64 pixels with 8 km of pixel size extracted from the NOAA/AVHRR images).

Furthermore, when 30 subjects who did not have prior knowledge about remote sensing were asked to use the proposed search system, the number of language hedges was too large.

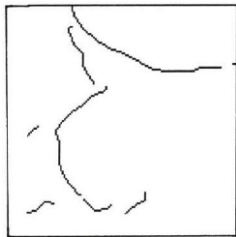


Fig. 6. Extracted spatial features (From the image of Fig.5, lines, edges, arcs and circles are extracted using the relaxation method with the thinning algorithm).

VI. CONCLUSION

A method for fuzzy retrievals of Earth observation satellite image database using geophysical parameters and spatial feature is proposed. It is confirmed that the proposed method allows fuzzy expressions of queries with sea surface temperature, chlorophyll-a concentration and cloud coverage as well as circle, line and edge, for instance “rather cold sea surface temperature and a sort of circle feature”. Thus users, in

particular, oceanographers may access the most appropriate image data from the database for finding of cold cores (circle features), fronts (arc and line features), etc. in a simple manner.

According to the retrieval method proposed in the paper, physical quantities and spatial features extracted from the earth observation satellite image to be retrieved are added to the inventory data, and retrieval can be performed faster and more efficiently using these. In addition, the search condition setting of those physical quantities and spatial features was allowed to be ambiguous, and the user's freedom of search was expanded. It also simplifies the method used by the user to enter conditions and displays browse images and their data as search results so that they can be easily viewed on the WWW, making it easy to visually confirm the results.

FUTURE RESEARCH WORKS

In the future, the proposed search method will be constructed as a database system and made public through WWW (<http://www.kyuic.gr.jp/>), and further improvement proposals from users will be made to be more practical.

ACKNOWLEDGMENT

The authors would like to thank to Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda for their valuable discussions.

REFERENCES

- [1] Ulrich Geske, Optimization and Simulation of Complex Industrial Systems, Proceedings of the 11th International Conference on Application of Prolog, INAP'98 (<http://www.oracle.com/>).
- [2] Isomoto, M., Nozaki, H., Yoshine, K., Nagai, H., Information Retrieval Techniques for Accumulated Data with High Fuzziness, Journal of the Japan Fuzzy Society, Vol. 8, No. 2, pp. 284-293, 1996.
- [3] Shinichi Sobue, Kohei Arai, Bunra Yoshida, Osamu Ochiai, Mina Ogawa, Mineo Sekiguchi, Tomotaka Sekiya, Masao Takagi, "A User-Friendly and Efficient Catalog Information Management and Provision Method for Earth Observation Satellite Data", Proceedings of Advanced Database System Proceedings of the Symposium'94 Symposium on Information Processing, Vol.94, No.13, pp.111-116, 1994.
- [4] Kohei Arai, Manabu Arakawa, Hirofumi Eto, Vague Search of Earth Observation Image Database Based on Fuzzy Theory Using Physical Quantities and Spatial Features, Journal of the Japan Society of Photogrammetry, Vol. 38, No. 4, pp. 17-25, Aug. 1999.
- [5] Kohei Arai, Hirofumi Eto, Tomoko Nishiyama, Earth Observation Satellite Image Database System Allowing Ambiguous Search Requests, Journal of the Japan Society of Photogrammetry, Vol.38, No.4, pp.47-52, Aug.1999 .
- [6] Shin-ichi Sobue, Kohei Arai, Fumiyooshi Yoshida, User Friendly and Efficient Catalog Information Management for Earth Observation Data, Information Processing Society of Japan, Advanced Database Workshop, pp. 111-116, December 1994, 1994
- [7] Kohei Arai, Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 32038, 2013.
- [8] Kohei Arai, Data collection and active database for tsunami warning system, Proceedings of the 1st International Workshop on Knowledge Cluster Systems, 2007.
- [9] Rosa Andrie, Achmad Basuki, Kohei Arai, A review of Chinese Academy of Science (CASIA) gait database as a human gait recognition dataset, Proceedings of the IES: Industrial Electronics Seminar, at EEPIS, 1-8, 2011.

² [https://en.wikipedia.org/wiki/Relaxation_\(iterative_method\)](https://en.wikipedia.org/wiki/Relaxation_(iterative_method))

- [10] Kohei Arai, R.Andrie, Gait recognition method based on wavelet transformation and its evaluation with Chinese Academy of Science (CASIA) gait database as human gait recognition dataset, Proceedings of the Information Technology for Next Generation: ITNG conference 2012, 213, 2012
- [11] Kohei Arai, Visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals, Journal of Visualization, DOI:10.1007/s, 12650-011-0118-6, 2011.
- [12] Kohei Arai, Visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals, Journal of Visualization, 15, 2, 155-166, 2012.
- [13] Kohei Arai, C.Rahmad, Wavelet based image retrieval method, International Journal of Advanced Computer Science and Applications, 3, 4, 6-11, 2012.
- [14] Kohei Arai, DP matching based image retrieval method with wavelet Multi Resolution Analysis: MRA which is robust against magnification of image size, International Journal of Research and Review on Computer Science, 3, 4, 1738-1743, 2012.
- [15] Kohei Arai, Free Open Source Software: FOSS based GIS for spatial retrievals of appropriate locations for ocean energy utilizing electric power generation plants, International Journal of Advanced Computer Science and Applications, 3, 9, 95-99, 2012.
- [16] Kohei Arai, Error analysis of air temperature profile retrievals with microwave sounder data based on minimization of covariance matrix of estimation error, International Journal of Advanced Computer Science and Applications, 3, 9, 85-89, 2012.
- [17] Kohei Arai, Visualization of link structure and URL retrievals utilization of interval structure of URLs based on brunch and bound algorithms, International Journal of Advanced Research in Artificial Intelligence, 1, 8, 12-16, 2012.
- [18] Kohei Arai, Method for image portion retrieval and display for comparatively large scale of imagery data onto relatively small size of screen which is suitable to block coding of image data compression, International Journal of Advanced Computer Science and Applications, 4, 2, 218-222, 2013.
- [19] Kohei Arai, Cahya Rahmad, Content based image retrieval by using multi-layer centroid contour distance, International Journal of Advanced Research in Artificial Intelligence, 2, 3, 16-20, 2013.
- [20] Kohei Arai, Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data, International Journal of Advanced Research in Artificial Intelligence, 2, 5, 32038, 2013.
- [21] Kohei Arai, Image retrieval and classification method based on Euclidian distance between normalized features including wavelet descriptor, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 19-25, 2013.
- [22] Kohei Arai, Numerical representation of web sites of remote sensing satellite data providers and its application to knowledge based information retrievals with natural language, International Journal of Advanced Research in Artificial Intelligence, 2, 10, 26-31, 2013.
- [23] Kohei Arai, Indra Nugraha Abudullar, Hiroschi Okumura, Image retrieval based on color, shape and texture for ornamental leaf with medicinal functionality, International journal of Image, Graphics and Signal Processing, Vol.6, No.7, June 2014
- [24] Cahya Rahmad, Kohei Arai, Comparison contour extraction based on layered structure and Fourier descriptor on image retrieval, International Journal of Advanced Computer Science and Applications, 6, 12, 71-74, 2015.
- [25] Kohei Arai, Pursuit Reinforcement Competitive Learning: PRCL Based Online Clustering with Tracking Algorithm and Its Application to Image Retrieval, International Journal of Advanced Research on Artificial Intelligence, 5, 9, 9-16, 2016.
- [26] Kohei Arai, Cahya Rahmad, Image Retrieval Method Utilizing Texture Information Derived from Discrete Wavelet Transformation Together with Color Information, International Journal of Advanced Research on Artificial Intelligence, 5, 10, 1-6, 2016.
- [27] S.Sobue and Kohei Arai, Metadata Definition and Retrieval of Earth Observation Satellite Data, Proceedings of the IEEE Metadata Conference, 1997.
- [28] Kohei Arai, Open GIS with spatial and temporal retrievals as well as assimilation functionality, Proceedings of the Asia Pacific Advanced Network Natural Resource Workshop, Utilization of Earth Observation Satellite-Digital Asia Special Session 1,p8,2003.
- [29] Kohei Arai, Geographic information system: GIS based on neural network for appropriate parameter estimation of geophysical retrieval equations with satellite remote sensing data, Proceedings of the IEEE Geoscience and Remote Sensing, PID 220128, 2006.
- [30] Kohei Arai, Yuji Yamada, Image retrieval method based on hue information and wavelet description based shape information as well as texture information of the objects extracted with dyadic wavelet transformation, Proceedings of the 11th Asian Symposium on Visualization, ASV-11-08-10, 1-8, 2011.
- [31] Kohei Arai, C.Rahmad, Wavelet based image retrievals, Proceedings of the 260th conference in Saga of Image and Electronics Engineering Society of Japan, 243-247, 2012.
- [32] Kohei Arai, Image Retrieval Method Based on Back-Projection, Proceedings of the Computer Vision Conference 2019.
- [33] Kohei Arai, Tran Xuang Sang, Fuzzy Genetic Algorithm for prioritization determination with techniques for order performance by similarity to ideal solution, International Journal of Computer Science and Network Security, 11, 5, 1-7, 2011.
- [34] Kamil Fagih, Wahyu Primadi, Anik Nur Harayani, Ari Priharta, Kohei Arai, Smart grid photovoltaic system pilot scale using sunlight intensity and state of charge (SoC) battery based on Mamdani fuzzy logic control, Journal of Mechatronics, Electrical Power and Vehicular Technology, 10, 36-47, 2019.
- [35] Kohei Arai, Satellite Image Database Search Engine Which Allows Fuzzy Expression of Geophysical Parameters of Queries, International Journal of Advanced Computer Science and Applications IJACSA, 11, 5, 69-73, 2020.
- [36] Luffi Mahardika, Anik Nur Hendayani, Heru Wahyn Herwanto, Kohei Arai, Operation of light tracker movement using Fuzzy logic control information and communication technology, Proceedings of the International Conference on IACT (ICOIACT 2018), Yogyakarta, 3D Parallel Session 3-D, 2018.

AUTHOR'S PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Science Commission "A" of ICSU/COSPAR during 2008 and 2020 then he is now award committee member of ICSU/COSPAR. He is now Visiting Professor of Nishi-Kyushu University since 2021, and is Visiting Professor of Kurume Institute of Technology (Applied AI Laboratory) since 2021. He wrote 87 books and published 700 journal papers as well as 570 conference papers. He received 66 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA.



<http://teagis.ip.is.saga-u.ac.jp/index.html>

Virtual Route Guide Chatbot Based on Random Forest Classifier

Puspa Miladin Nuraida Safitri A. Basid¹, Fajar Rohman Hariri², Fresy Nugroho³,
Ajib Hanani⁴, Firman Jati Pamungkas⁵

Informatics Engineering, Faculty of Science and Technology,
Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia^{1,2,3,4}
Library and Information Science, Faculty of Science and Technology,
Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia⁵

Abstract—Improvements in the quality of tourism services and the number of human resources will affect the quality of social services and information services provided to foreign tourists, thereby enhancing the quality of services offered regarding tourist destination information in the Malang Raya area. Considering the urgency of foreign tourists in obtaining information related to directions, routes, and access roads to their desired tourist destinations, especially in East Java, due to limited data from the government agencies handling the tourism sector, as well as the difficulty in communication with residents who may not understand what is being communicated by foreign tourists. Therefore, the need for an interactive chatbot to assist in obtaining routes and access information to the desired tourist destinations will facilitate foreign tourists. To improve the accuracy of the chatbot's ability to answer sentence selection, the use of artificial intelligence, specifically the Random Forest Classifier, is necessary. This study obtained the highest accuracy value using a tree quantity of 200, a maximum tree depth of 20, and a minimum sample split of 5. Using these quantities resulted in an accuracy of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%.

Keywords—Tourism; chatbot; artificial intelligence; random forest classifier

I. INTRODUCTION

Tourism is a dynamic activity involving many people and stimulating various business sectors. In the current era of globalization, the tourism sector will become the main driver of the world economy and global industry. Tourism will provide significant revenue for regions aware of its potential in the tourism sector [1]. Thus, tourism has become an integral part of human life. The tourism sector has led several regions in Indonesia to develop tourism as their distinctive feature.

One area with great tourism potential is Malang, located in the East Java province of Indonesia. Malang Raya has three administrative areas: Malang Regency, Malang City, and Batu City. Malang Raya is one of the leading tourist destinations in East Java and Indonesia. According to statistical data compiled by the Ministry of Tourism and Creative Economy in 2022, the total number of international tourists visiting Malang Raya reached six million [2].

The Malang Raya region offers many beautiful tourist destinations unique to other areas in Indonesia. Malang Raya offers various tourism categories, from natural to

manufactured attractions. According to data from the Malang Regency's One Data program, there are 16 village tourism objects, 106 natural tourist attractions, 49 cultural tourist attractions, and 24 manufactured tourist attractions in the Malang Regency area [3]. According to data from the Central Statistics Agency (BPS) of Batu City, there are ten manufactured tourist attractions, 12 village tourism objects, five natural tourist attractions, five souvenir tourist attractions, and one religious tourist attraction in the Batu City area [4]. According to data from Malang's One Data program, there are 16 cultural tourist attractions, two historical tourist attractions, four religious tourist attractions, one educational tourist attraction, 2015 culinary tourist attractions, 12 shopping tourist attractions, and 20 manufactured tourist attractions in the Malang City area [5].

The abundance of tourist destinations in the Malang Raya region provides many options for international tourists. However, on the other hand, with the increasing number of tourist destinations, issues arise regarding information about the destinations to be visited. International tourists require comprehensive information about routes, directions, and ways to guide their travel, but not all available information from print media, television, the Internet, and other sources can meet these needs [6]. Another issue in the tourism sector is the low quality of service and quantity of human resources in the tourism industry worldwide. This aspect needs special attention in efforts to improve the tourism sector in the Malang Raya region. The quality of tourism services and the quantity of human resources are among the standards that will be compared to achieve tourist satisfaction [7]. Improvements in the quality of tourism services and the number of human resources will affect the quality of social service and information services provided to international tourists, thereby enhancing the quality of services related to tourist destination information in the Malang Raya region.

Another issue encountered is the current condition of the official websites of the Department of Culture and Tourism of Malang Regency, Malang City, and Batu City, which only provide brief information about tourist attractions and lacks interactive question-and-answer features that can guide foreign tourists in terms of directions, routes, and roads to the destinations. As a result, foreign tourists have to search for routes or access on their own through several stages. The lack of digital information services like this can lead to inefficiency

and ineffectiveness in obtaining access information or routes to selected tourist destinations by foreign tourists [7].

Another issue is the poor communication between residents and foreign tourists. This is due to the lack of knowledge, which prevents residents from understanding what foreign tourists communicate [8]. This has become a significant concern for the government, notably the Tourism Sector Institution in the Malang Raya area. Considering the urgency and the difficulty faced by foreign tourists in obtaining information regarding directions, routes, and access to tourist destinations, especially in East Java, due to limited information from the government agencies handling the tourism sector, as well as the difficulty in communication with residents who do not all understand what foreign tourists are communicating. According to the author, there is a need for an interactive chatbot that can assist in obtaining route information and access to the desired tourist destinations, which will facilitate foreign tourists.

Artificial intelligence must be utilized to improve the chatbot's accuracy in answering sentence selection. Several previous studies have added artificial intelligence to building chatbots. One of which is creating a chatbot related to Covid-19 [9]. Other research is also being carried out in building chatbots with artificial intelligence, namely a classification method to identify intentions rather than user input, it is called purpose classification in the chatbot system [10]. Furthermore, in the field of tourism, another research has been conducted by creating a website along with a chatbot for the city of Kanazawa [11]. The method used in this research is the Random Forest Classifier method. Random Forest Classifier is an algorithm that results from the bootstrapping aggregation of Decision Tree algorithms. This research employs this method due to its advantages over other algorithms, as it falls into Classification and Regression Tree (CART) methods, which utilize historical data to build a decision tree.

Based on the background, the author believes that an interactive chatbot capable of assisting foreign visitors in obtaining information about routes and ways to reach their desired tourist destinations will greatly facilitate them. In this study, it is expected that an interactive chatbot using the Random Forest Classifier method can optimize the accuracy level in sentence prediction performance and utilize the Telegram Messenger to structure the data more effectively while also providing social services.

II. LITERATURE REVIEW

A. Chatbot

Chatbot is a program in artificial intelligence designed to communicate directly with humans as its users. The difference between a chatbot and a natural language processing system is the algorithms' simplicity. Although many bots can interpret and respond to human input, they only interpret keywords in the input and reply with the most suitable keywords or patterns of words from pre-existing data in a database created beforehand [12]. The future of Software Engineering is expected to undergo a significant transformation with the emergence of chatbots. These chatbots will enable software practitioners to communicate and inquire about their projects

using everyday language, revolutionizing how they interact with various services. At the core of every chatbot lies a Natural Language Understanding (NLU) component, which empowers the chatbot to comprehend and interpret human language inputs [13].

Initially, these computer programs (bots) were tested through the Turing test, which involved concealing their identity as machines to deceive the person conversing with them. If a user cannot identify the bot as a computer program, that chatbot is categorized as artificial intelligence.

One famous chatbot is Eliza (Dr. Eliza), developed by Joseph Weizenbaum at the Massachusetts Institute of Technology (MIT). Eliza is a pioneering chatbot known as a chat program that plays the role of a psychiatrist. Eliza simulates conversations between a psychiatrist and their patients in natural English. Eliza was created to study natural language communication between humans and machines. Eliza acts as a psychologist who can answer the patient's questions with reasonable responses or respond with further questions [12].

B. Random Forest Classifier

Random Forest is a method introduced by Breiman, a development and combination of multiple Decision Trees. While a Decision Tree represents a single classification tree, Random Forest creates multiple trees to determine its prediction results. Combining bootstrap aggregating and random feature selection in a Random Forest can reduce the overfitting problem in small training data [14]. Since Random Forest is an ensemble method of CART, it does not assume or work well in non-parametric cases.

The steps involved in the Random Forest as shown in Fig. 1 are explained as follows:

- Determining the parameters of Random Forest. The value of $mtry$, or the number of randomly selected predictor variables, is set to $\frac{p}{3}$ for regression cases, where p is the total number of predictor variables [14].
- Then, determine the recommended number of N_{tree} trees to use, typically 50 trees. According to Breiman [14], 50 trees provide satisfactory results for classification cases. However, $N_{tree} \geq 100$ yields lower misclassification rates [15].
- Specifying the stopping criteria default in scikit-learn Random Forest, where one means that if a subnode/child node contains only 1 sample, it will stop splitting and become a terminal node/leaf node. Therefore, once the branching stops, terminal nodes are generated as the prediction result of a single CART tree [16].
- Splitting the data into training and testing datasets. From the training data, n samples are randomly selected with replacement (bootstrap) to create a new dataset D , where I represents the bootstrap sample division of the i -tree.

- The bootstrap sampling only takes $\frac{1}{3}$ of the entire training data and the remaining $\frac{2}{3}$ is considered out-of-bag (OOB) data, which is useful for measuring the performance of regression trees [17].
- Making predictions based on constructing tree models from the new dataset D, using a combination of randomly selected m predictor variables (random feature selection).

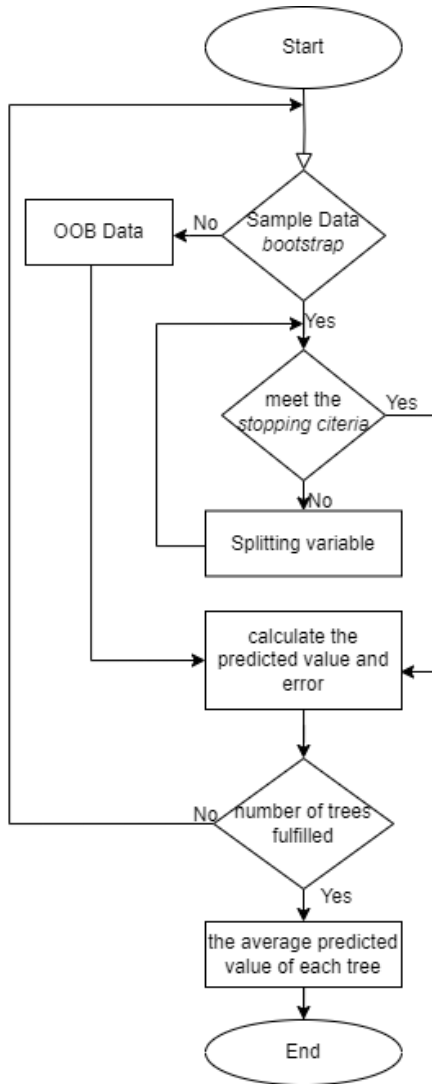


Fig. 1. Flow diagram of random forest.

III. PROPOSED MODEL

The initial stage involves data collection. Data collection for this research uses references from Google, specifically collecting frequently used conversational sentences in English. The data collection for this research is divided into two parts: training data collection and testing data collection. While collecting training data, the author used 100 conversational sentences in English. These sentences were then labeled based on their sentence classes, whether they belonged to the statement (S), question (Q), or chat (C) class. Table I provides

an example of the conversational sentences used as training data, labeled based on their sentence classes.

TABLE I. EXAMPLE DATASET OF SENTENCE FOR MODEL FORMATION AND CLASSIFICATION

No	Sentence	Label
1	Sorry, I don't know about the weather.	S
2	What's the weather like today?	C
3	I am fine	C
4	Where do you live?	Q
5	Are you a chatbot?	Q

After labeling the dataset consisting of 100 conversational sentences, the dataset will be processed by assigning Part-of-Speech (POS) tags to each word in the corpus. Fig. 2 displays the type of Penn treebank tagset. Once POS tags are assigned, the patterns for classifying the training sentence models will be determined. Part-of-Speech (POS) tagging, or simply tagging, is the process of assigning syntactic labels or Part-of-Speech tags to each word in the corpus. Since tags are generally applied to punctuation marks, punctuation marks such as periods, commas, etc., must be separated from the words during the tagging process. The extracted features from the data are used to build the required model by extracting Part-of-Speech (POS) tags, resulting in numerical data features. Fig. 3 provides an example of the POS tagging process in this research.

After that, the tags in each sentence will be calculated according to the predetermined tags, serving as a reference for pattern formation within the sentences. The tags used as markers in this process include a cardinal number (CD), noun singular or mass (NN), proper noun singular (NNP), proper noun plural (NNPS), noun plural (NNS), personal pronoun (PRP), verb gerund or present participle (VBG), and verb 3rd person singular present (VBZ). Table II provides an example dataset table for counting the number of Part-of-Speech (POS) tags.

Tag	Description	Example	Tag	Description	Example
CC	coord. conjunction	and, or	RB	adverb	extremely
CD	cardinal number	one, two	RBR	adverb, comparative	never
DT	determiner	a, the	RBS	adverb, superlative	fastest
EX	existential there	there	RP	particle	up, off
FW	foreign word	noire	SYM	symbol	+, %
IN	preposition or sub-conjunction	of, in	TO	"to"	to
JJ	adjective	small	UH	interjection	oops, oh
JJR	adject., comparative	smaller	VB	verb, base form	fly
JJS	adject., superlative	smallest	VBD	verb, past tense	flew
LS	list item marker	1, one	VBG	verb, gerund	flying
MD	modal	can, could	VBN	verb, past participle	flown
NN	noun, singular or mass	dog	VBP	verb, non-3sg pres	fly
NNS	noun, plural	dogs	VBZ	verb, 3sg pres	flies
NNP	proper noun, sing.	London	WDT	wh-determiner	which, that
NNPS	proper noun, plural	Azores	WP	wh-pronoun	who, what
PDT	predeterminer	both, lot of	WP\$	possessive wh-	whose
POS	possessive ending	's	WRB	wh-adverb	where, how
PRP	personal pronoun	he, she			

Fig. 2. Penn treebank tagset.

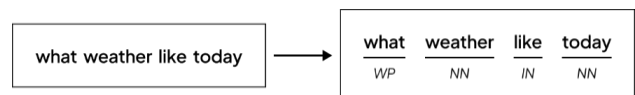


Fig. 3. Example of POS tagging.

TABLE II. EXAMPLE DATASET FOR COUNTING THE NUMBER OF PART-OF-SPEECH (POS) TAGS

No	Word Totals	C	N	NNP	NNS	PRP	VBG	VBZ
1	7	0	1	1	0	0	1	0
2	5	0	2	0	0	0	0	1
3	3	0	0	0	0	0	1	0
4	4	0	0	0	0	0	1	0
5	4	0	1	0	0	0	1	0

After collecting the training dataset of conversational sentences and the Part-of-Speech patterns for building the classification model, the next step is to determine the dataset of conversational sentences to be used as the output sentences for the chatbot based on the formed training classification model. The dataset represents the data used in the testing process. There are 4,085 conversational sentences that will be used as the dataset in the testing process. Subsequently, this data will be labeled according to the sentence classes. The labeling is based on the criteria of whether the sentence belongs to the question (Q) or chat (C) class, similar to the labeling in the first dataset. Table III provides an example dataset table of conversational sentences that will be used in the testing process.

TABLE III. EXAMPLES OF OUTPUT SENTENCE DATASETS USED IN THE TESTING PROCESS

No	Sentence	Label
1	Hi there, how are you!?	C
2	My name is Tourist Chatbot, but you can call me Lisa	C
3	ok, thanks!	C
4	Do you like it?	Q
5	What's that?	Q

The next stage is to process the user's input in conversational sentences. The entered conversational sentences by the user will go through the processing stage, which includes preprocessing and training processes. In the preprocessing stage, a Natural Language Processing (NLP) approach is used, which involves case folding to convert all letters in the document to lowercase, tokenization to separate input words into individual tokens, stemming from finding the base form of words and producing the correct language structure, and POS tagging to assign Part-of-Speech tags or syntactic classes to each word in the corpus.

After the preprocessing process, the data will proceed to the training process, which involves training the preprocessed data using the Random Forest Classifier approach (see Fig. 4). In this data training stage, a classification model will be formed through the training using the Random Forest Classifier method. The final stage is testing, which is conducted to obtain accuracy and error values for the chatbot. Once the process is completed, the system will generate an output sentence to be sent as a response to the user's input sentence.

IV. IMPLEMENTATION

The implementation phase is the stage of applying the designed system based on the system design that has been created. The results of this system implementation are used to classify sentences in the virtual route guide chatbot using the random forest classifier method. The implementation phase begins by inputting data into the system that has been created using the Python programming language. The labeled sentence data, stored in .csv format, is inputted into the system, and the column containing the conversational sentences and sentence classes is extracted.

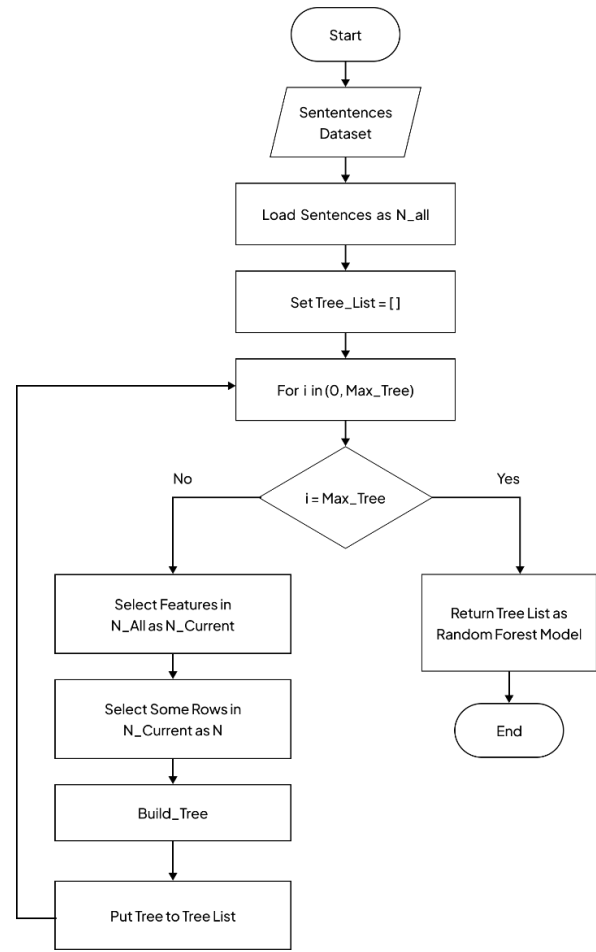


Fig. 4. Random forest classifier algorithm flowchart.

Next, the system proceeds to the initialization stage of part-of-speech patterns for sentence formation, which will be used in feature selection. Several part-of-speech patterns are used in this initialization stage, such as VerbCombos, questionTriples, statementTriples, startTuples, and endTuples. Then, feature_keys are initialized as keywords for the part-of-speech tags used.

After completing the initialization stage, the system moves to the preprocessing stage, including case folding, tokenizing, stemming, and POS tagging. The case folding process utilizes the lower() function to convert the entire sentence to lowercase to avoid case sensitivity. The tokenizing process uses the split() function to separate the sentence into individual words. The stemming process utilizes libraries from

NLTK with the WordNetLemmatizer and SnowBall algorithms. The POS tagging process assigns Part-of-Speech tags to each word in the corpus.

The next step is to create a function to count the frequency of occurrence of part-of-speech tags in each sentence. The processed data, using the function to count the frequency of part-of-speech tags, is then stored in a data frame containing the calculated frequency counts and saved in a new CSV file. This data calculates the number of part-of-speech tags in each sentence. Once the calculated frequency counts of part-of-speech tags are successfully saved, the dataset is imported and loaded into Python. Then, the class variable with the object data type is converted to the int64 data type. This is because the class variable represents a category that indicates it is categorically ordinal.

Next, the feature-selected data stored in a new file will be divided into training and testing data. The training data will be implemented into the system that has been created to build a sentence classification model using the random forest classifier. Based on the rule of thumb, the training and testing data will be split in a 75:25 ratio, with 4,185 pattern data points to be implemented in the system. This data split is done using the model_selection library's train_test_split function.

After dividing the data into training and testing sets, the next step is to train the dataset by creating a classification model using the random forest classifier. The aim is to classify data effectively and efficiently. This model is used to classify data into multiple categories or classes with high accuracy. In creating this classification model, the feature extraction function is used to improve the accuracy of the classification model by eliminating irrelevant features and retaining the most important features in the data. By selecting the most relevant and important features, the classification model can find more accurate patterns and distinguish between different classes better.

Once the training model is created, the next step is to create a Telegram account and access @BotFather. @BotFather is the official Telegram bot that allows users to create their own Telegram bot. After creating the Telegram bot, the next step is to install the python-telegram-bot library. This library is used to connect the bot to the Telegram API and send and receive user messages.

The next step is to write Python code for the bot and add message-handling functions that the bot will process. In creating the Python code, the telegram.ext module from the python-telegram-bot library is required. Then, an Updater object is created with the bot token and the bot is started using the command "updater.start_polling()". The message-handling function will be executed every time the bot receives a new message from the user. Specific logic is added to the message handling function to process user requests and provide appropriate responses based on the received message.

V. EXPERIMENTAL RESULT

In this research, there are two datasets, each consisting of 100 and 4,185 labeled conversational sentences according to the rules of Part-of-Speech (POS) tagging, based on the criteria of whether the sentences belong to the question (Q) or chat (C) class. Table IV shows the number of sentences in each label within the conversational sentence dataset used as training data for the classification model.

TABLE IV. NUMBER OF SENTENCES PER LABEL

Dataset 1		
No.	Label	Number of Sentences
1	S (Statement)	32
2	Q (Question)	43
3	C (Chat)	25
Total Number of Sentences		100
Dataset 2		
No.	Label	Number of Sentences
1	S (Statement)	1473
2	Q (Question)	1.238
3	C (Chat)	3.073
Total Number of Sentences		4.185

The data will be divided into two parts: training and testing data. The training data will be implemented into the system that has been created to build a classification model using the random forest classifier for sentences. Meanwhile, the testing data is used to evaluate the system's performance. The conversational sentence dataset will be divided into testing data and training data with a ratio of 75:25 based on the rule of thumb. In addition, experiments are conducted by varying the number of trees (n_estimators), tree depth (max_depth), and the minimum number of samples required to split a node (min_sample_split). The number of trees tested is 100, 200, and 500, the depth of the trees tested is 5, 10, and 20, and the minimum samples required to split a node tested are 2, 5, and 10.

From Table V, the highest accuracy value is obtained using 200 trees, with a maximum tree depth of 20 and a minimum sample split of 5. Using these parameters yields an accuracy value of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%. On the other hand, the lowest accuracy value is obtained from using 500 trees, with a maximum tree depth of 5 and a minimum sample split of 10. Using these parameters yields an accuracy value of 94.43%, precision of 95.29%, recall of 94.61%, and f-measure of 94.95%. These accuracy results are also depicted in the visualization graph, as shown in Fig. 5.

TABLE V. SHOWS THE TESTING RESULTS USING 100, 200, AND 500 TREES WITH A COMBINATION OF 20 MINIMUM SAMPLES AND 10 MINIMUM SAMPLES REQUIRED TO SPLIT A NODE

Tree	Deep Max. Tree	Min. Sample	Accuracy	Precision	Recall	F-measure
100	5	2	94,58%	95,30%	94,80%	95,05%
		5	94,73%	95,40%	94,95%	95,17%
		10	94,66%	95,35%	94,87%	95,11%
	10	2	95,42%	95,94%	95,61%	95,78%
		5	95,34%	95,89%	95,53%	95,71%
		10	95,11%	95,70%	95,32%	95,51%
	20	2	95,50%	95,83%	95,73%	95,78%
		5	95,73%	96,12%	95,91%	96,02%
		10	95,50%	96,00%	95,65%	95,82%
200	5	2	94,66%	95,45%	94,83%	95,14%
		5	94,73%	95,50%	94,91%	95,21%
		10	94,58%	95,40%	94,76%	95,08%
	10	2	95,57%	96,08%	95,75%	95,92%
		5	95,50%	96,07%	95,65%	95,86%
		10	95,42%	96,01%	95,57%	95,79%
	20	2	95,42%	95,74%	95,71%	95,72%
		5	95,88%	96,29%	96,03%	96,16%
		10	95,57%	96,05%	95,73%	95,89%
500	5	2	94,50%	95,35%	94,69%	95,02%
		5	94,50%	95,35%	94,69%	95,02%
		10	94,43%	95,29%	94,61%	94,95%
	10	2	95,50%	96,00%	95,69%	95,84%
		5	95,42%	95,98%	95,59%	95,78%
		10	95,42%	95,98%	95,59%	95,78%
	20	2	95,50%	95,76%	95,82%	95,79%
		5	95,80%	96,24%	95,95%	96,09%
		10	95,57%	96,05%	95,73%	95,89%

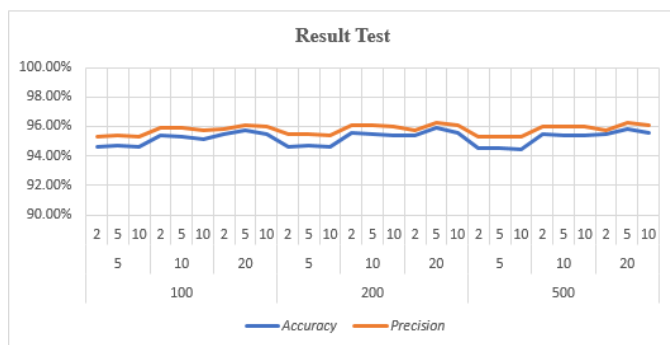


Fig. 5. Visualization of test results in conversational sentences.

VI. CONCLUSION

The result of this research is an interactive chatbot created using the Random Forest Classifier method, which can optimize the level of accuracy in sentence prediction performance and utilizes Telegram Messenger to structure the

data more effectively. According to the test results in the previous section, this study obtained the highest accuracy value using a tree quantity of 200, a maximum tree depth of 20, and a minimum sample split of 5. Using these quantities resulted in an accuracy of 95.88%, precision of 96.29%, recall of 96.03%, and f-measure of 96.16%. This proves that the use of random forest classification affects the sentence classification results. The reason is the existence of feature selection that can reduce the features used.

ACKNOWLEDGMENT

We express our deepest gratitude to all members of the Informatic Engineering, Faculty of Science and Technology, UIN Maulana Malik Ibrahim, Malang.

REFERENCES

- [1] Ismayanti, Introduction to Tourism. Jakarta: Grasindo, 2010.
- [2] Tugu Malang, "6 million tourists are visiting Malang, the number exceeds the 2022 target," Tugu Malang, 2022.

- [3] Communication and Informatics Office of Malang Regency, KABUPATEN MALANG SATU DATA. 2022.
- [4] Statistics of Malang Municipality, "Number of Domestic Tourists in Malang City (People) in 2020-2022," <https://malangkota.bps.go.id/indikator/16/157/1/jumlah-wisatawan-domestik-di-kota-malang.html>, 2022.
- [5] Statistics Batu City, "Number of Visitors to Tourist Attractions and Tourism Souvenirs According to Tourist Attractions in Batu City in 2021." [Online]. Available: <https://batukota.bps.go.id/statictable/2022/04/11/1383/jumlah-pengunjung-objek-wisata-dan-wisata-oleh-oleh-menurut-tempat-wisata-di-kota-batu-2021.html>. [Accessed: 26-Apr-2023].
- [6] M. Aziz and M. Aman, "Decision Support System For Selection Of Expertise Using Analytical Hierarchy Process Method," 2019.
- [7] D. Rinova, "Proceedings Sustainable Development Goals (SDGs) Conference International Science Consortium for Indonesian Sustainability (ISCIS) Analysis of Tourist Attraction and Service Quality on Tourist Satisfaction."
- [8] M. A. Cholik, "THE DEVELOPMENT OF TOURISM INDUSTRY IN INDONESIA : CURRENT PROBLEMS AND CHALLENGES," *Eur. J. Res. Reflect. Manag. Sci.*, vol. 5, no. 1, 2017.
- [9] W. Astuti, D. P. I. Putri, A. P. Wibawa, Y. Salim, Purnawansyah, and A. Ghosh, "Predicting Frequently Asked Questions (FAQs) on the COVID-19 Chatbot using the DIET Classifier," 3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021, pp. 25–29, 2021, doi: 10.1109/EIConCIT50028.2021.9431913.
- [10] M. Y. H. Setyawan, R. M. Awangga, and S. R. Efendi, "Comparison Of Multinomial Naive Bayes Algorithm And Logistic Regression For Intent Classification In Chatbot," *Proc. 2018 Int. Conf. Appl. Eng. ICAE 2018*, pp. 1–5, 2018, doi: 10.1109/INCAE.2018.8579372.
- [11] D. Suzuki, K. Nunotani, K. Fukusato, and M. S. Tanaka, "A Study of Tourism Proposal System Using AI," 2020 IEEE 9th Glob. Conf. Consum. Electron. GCCE 2020, pp. 634–635, 2020, doi: 10.1109/GCCE50665.2020.9292070.
- [12] J. Weiznbaum, "ELIZA - A Computer Program for the Study of Natural LanguageCommunication Between ManAnd Machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1996.
- [13] A. Abdellatif, K. Badran, D. E. Costa, and E. Shihab, "A Comparison of Natural Language Understanding Platforms for Chatbots in Software Engineering," *IEEE Trans. Softw. Eng.*, vol. 48, no. 8, pp. 3087–3102, 2022, doi: 10.1109/TSE.2021.3078384.
- [14] L. Breiman, J. Friedman, C. J. Stone, and R. . Olshen, *Classification and Regression Trees*, 1st ed. Boca Raton : Taylor & Francis Group, 1984.
- [15] C. D. Sutton, "Classification and Regression Trees, Bagging, and Boosting," *Handbook of Statistics*, vol. 24. Elsevier, pp. 303–329, 2005, doi: 10.1016/S0169-7161(04)24011-1.
- [16] F. Widmaier, A. Zell, A. Schilling, and J. Bohg, "Robot Arm Tracking with Random Decision Forests."
- [17] S. Liu et al., "Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization," *Comput. Electron. Agric.*, vol. 95, pp. 82–91, 2013, doi: 10.1016/j.compag.2013.03.009.

Approaches and Tools for Quality Assurance in Distance Learning: State-of-play

Silvia Gaftandzhieva¹, Rositsa Doneva², Senthil Kumar Jagatheesaperumal³

Faculty of Mathematics and Informatics, University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria¹

Faculty of Physics and Technology, University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria²

Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India³

Abstract—In recent years, distance learning has become an increasingly popular mode of education due to its flexibility and accessibility. However, the quality of distance learning programs has been a cause for concern, which has led to the development of various approaches and tools for quality assurance and assessment. This review article aims to provide an in-depth analysis of the current state of play of quality assurance in distance learning. The paper discusses the fundamental requirements to establish quality in distance learning and the challenges associated with ensuring quality in this mode of education. Then it explores the different approaches and tools used for quality assurance and assessment, such as course evaluations, self-assessments, and external reviews. In addition, the paper delves into the development of regulatory documents and manuals for quality assurance, which are essential for ensuring that distance learning programs adhere to established standards. It also discusses in detail the importance of audits and accreditations from assessment organizations in assuring quality in distance learning. As the satisfaction of all stakeholders (including students, faculty, and administrators) is crucial for ensuring the success of distance learning programmes, the paper highlights the various measures HEIs can take to ensure stakeholder satisfaction. Finally, the article discusses the processing of statistical data and performance indicators, which can provide valuable insights into the effectiveness of distance learning programmes.

Keywords—Distance learning; quality assurance; assessment; stakeholder satisfaction; regulatory documents; performance indicators

I. INTRODUCTION

The rapid adoption of distance education in higher education institutions (HEIs) worldwide has brought about demands from scientific and technological developments and some challenges to quality, including technological problems, administration, instructional methods, and student barriers. Quality management in distance learning courses is vital for improving the overall educational experience, yet the indeterminate definitions of quality make it challenging to evaluate effectively [1]. Students and educators face barriers such as low self-organization, lack of effective interaction, and a sense of isolation that can decrease their satisfaction with online learning. While distance learning courses have prevented transfer from theory to design practices, using a quality assurance model for web-based learning and implementing benchmarks[2], e.g. in course development, teaching/learning process, course structure, and faculty support categories, can help. The six dimensions for measuring service

quality in distance education are tangibles, reliability, responsiveness, delivery, assurance, and student participation [3].

The shift towards distance learning has also raised concerns about the impact on crucial social learning aspects, particularly for early elementary children and vulnerable student populations. The move to a distanced setting has reduced opportunities for face-to-face interaction, which is essential for promoting collaboration, teamwork, and socialization. However, the use of innovative technology, such as virtual classrooms, online discussion forums, and social media platforms, can provide alternative ways for students to interact and engage with their peers and instructors [4]. Furthermore, the quality of distance learning depends only on the course design and delivery but also on the level of readiness and support provided by the government and educational institutions. There is a significant association between parents' satisfaction with the quality of education, how they assess teachers' competencies and the level of government readiness to switch to a distance learning format [5]. Therefore, it is essential HEIs to ensure adequate resources and infrastructure and that instructors are adequately trained and equipped to deliver quality distance education.

Another concern is while distance learning cannot replace traditional education to a full degree, it can serve as a valuable complement. HEIs can utilize distance learning to enhance the knowledge and skills of students in various areas [6], such as emerging technologies, digital literacy, and educational technology. Moreover, the relationship between teachers and students has been identified as a significant factor in determining students' satisfaction with distance learning courses. A positive teacher-student relationship plays an intermediary role when linking the attitudes and behaviour of teachers with students' overall satisfaction with learning courses [7].

Additionally, distance learning cannot fully substitute traditional learning forms in HEIs, but distance learning can provide valuable enhancements. It is crucial to acknowledge that transitioning to distance learning may limit certain social learning aspects, particularly among younger students and vulnerable populations. However, HEIs can leverage distance learning to reduce costs and attain sustainable advantages [8]. Furthermore, the relationship between the attitudes and behaviour of teachers and students' overall satisfaction with their courses should be examined through mediation,

underscoring the critical success factors within the quality assurance framework.

Although the topic of quality assurance of distance education has been widely studied by researchers for decades, there is a need to systematize all possible ways to ensure the quality of education.

This paper explores the essential components of quality management in distance education. Section II delves into the specific requirements needed to ensure quality in distance learning. It examines the fundamental aspects that must be considered and implemented to maintain high standards in delivering educational content. Section III focuses on the various definitions of quality in the context of distance learning. As reports of the diverse perspectives and interpretations of quality, this section provides clarity and understanding of the concept, laying the foundation for efficient evaluation. The paper then discusses approaches and tools for assuring and assessing quality in distance education. These include the development of normative documents and manuals that serve as guidelines and benchmarks for quality assurance. In addition, the section emphasizes the importance of conducting audits and accreditation by reputable assessment organizations. The paper further highlights the significance of gathering feedback through periodic surveys among stakeholders, including academic staff, students, and external entities. Finally, it explores the utilization of intelligent data analysis tools, enabling deep insights into the effectiveness and quality of distance learning programs. The Conclusion summarizes the contributions and plans for future work.

II. QUALITY OF DISTANCE LEARNING: DEFINITIONS

Researchers and various stakeholders (students, teachers, HEIs leadership, employers, external evaluators, etc.) widely discuss issues related to the quality of distance learning. All emphasize the need for a better understanding of the aspects that contribute to achieving high quality in distance learning programs [9]. There is no universally accepted definition of quality in distance learning. Quality is a multidimensional concept encompassing a wide range of products, services, supplies and philosophies and attempts to meet the needs and expectations of students and various stakeholder groups with different interests [10].

According to Robinson [11], the quality of distance learning can be the result of various factors, both internal and external to the HEI, for example, the skill levels and experience of the staff, the number of resources available, weak or strong leadership, the effectiveness of administrative systems and the communication infrastructure.

Some researchers emphasize adherence to standards and procedures [10], that apply to each course and influence course design, course layout and the amount of learning content. Burns [12] defines quality as adherence to a set of standards for content, design, and instruction, and Lassoued, Alhendawi, and Bashitialshaer [13] - as a set of procedures and guidelines adopted in the educational institution that supports the management of the organization and provision of services. According to Roe [14], significant components for developing quality distance courses are the assurance of rich multimedia,

asynchronous communication, and faculty mentoring. Achieving quality teaching and learning is a complex endeavour involving multiple dimensions [15], including curriculum design and course content, learning contexts, use of feedback, assessment of learning outcomes, learning environment, and student support services.

Other researchers emphasize the design and delivery of training. To achieve a high quality of training, the academic staff must have experience in developing educational content for distance learning [16], skills in using technology, and applying modern pedagogical approaches to teaching and guiding students in the learning process [17-24]. In addition, educators should implement forms of communication and interaction that are student-centred and encourage their active participation in the learning process [25], provide support to students [26], and use forms of assessment consistent with individual or group distance learning approaches and stimulating critical thinking [16]. Elias [27] presents eight instructional design principles for the quality of distance learning courses - Equitable Use, Flexible Use, Simple and Intuitive, Perceptible Information, Tolerance for Error, Low Physical and Technical Effort, Community of Learners and Support, and Instructional climate. According to McClary [28], high-quality distance courses are those in which the learning content is up-to-date, each module contributes to a specific course objective, the instructor provides the necessary support to students, and there is an effective support system. Lee and Dziuban [29] believe that the success of e-learning largely depends on strategies for evaluating the quality of the distance learning program. Grutzner, Weibelzahl, and Waterson propose four dimensions for assessing the quality of e-courses [30]: the content of learning materials, presentation of learning materials, teaching style, and overall course functioning. According to them, HEIs must consider these dimensions simultaneously and continuously throughout the life cycle of learning e-courses to ensure a high-quality product. Clayton Wright sets out criteria for evaluating the quality of e-courses [31], including general course information, information accessibility, course organization, language, layout, goals and objectives, course content, learning strategies, practice opportunities, learning resources, and assessment.

A third group of researchers emphasize student outcomes [12] and argue that the quality of distance learning can be spoken of when there is evidence that students leave with relevant knowledge and skills for post-graduation employment and employer satisfaction. A significant factor in improving desired learning outcomes and student satisfaction is students' engagement during learning [32-33]. According to Markova, Glazkova, and Zaborova [1], student satisfaction is influenced directly by the skills of teachers to use active learning techniques effectively, integrating high-level interaction and collaboration in instructional design, and ensuring high-quality and timely support and resources for learners.

Another group of researchers defines more components for the quality of the learning process in distance learning. According to Zaman, Ghosh, Datta and Basu [34], the quality of e-learning depends on the quality of the learning content, the quality of the learning management system from a technological point of view (ease of use, reliability, etc.) and

the quality of services (support of e-learning participants). Montedoro and Infante indicate three dimensions of the quality of learning management systems [35]: Technology, Content and Services. Lanzilotti, Ardito and Costabile [36] define the quality of the learning management system as the extent to which the technology, interaction, content, and services offered match the expectations of teachers and learners and enable them to teach/learn with pleasure.

Opponents of distance learning outline issues related to the quality and effectiveness of distance education compared to conventional educational models caused by several reasons. The main one is the ever-increasing demand for informed human resources to participate effectively in the global market [37]. Proponents of distance learning [25], [38-39] argue that distance learning should be as effective as face-to-face learning. HEIs can prove this by demonstrating that the quality of content, delivery, assessment, and outcomes in distance learning is equal to or better than traditional forms of education [9].

E-learning and distance learning are growing in popularity during the COVID-19 pandemic. The imposed restrictions catalyzed the digital transformation and modernization of all educational processes [10], [40-43], including in universities without previous experience in organizing and conducting distance learning. The forced transition to online learning presents HEIs with the challenge of fully transitioning to online learning while maintaining the quality of the education provided [44]. It also increases the interest of researchers in developing tools for assessing the quality of the functioning of educational software in HEIs [10], [43] and identifying elements that influence student satisfaction and allowing HEIs to develop strategies to ensure the quality of digital transformation [42].

According to Robinson [11], an aspect that is receiving increasing attention is how HEIs manage their quality regardless of their structure, context, or circumstances. Quality assurance is an approach to quality management that focuses on process management and aims to demonstrate and improve the quality of educational products and outcomes, to enable systematic management and monitoring of performance against set objectives [11]. Implementing a quality assurance system in higher education requires a share of responsibilities between the managers and all stakeholders [45-49]. It implies solving challenging tasks to address multiple dimensions, aspects, and meanings of quality from different perspectives and interests [10]. The quality assurance system may include a well-defined set of principles and procedures for achieving the overall goals of the institution, standards of achievement, established ways of responding to problems and clear accountability for results, a plan for training and development of staff, monitoring procedures of performance [10-11], [38]. Because HEIs must make continuous efforts to exceed the expectations of students and stakeholders, a good quality assurance system must be focused on student satisfaction [10] and periodically updated. As a result of implementing a systematic and consistent quality assurance system, HEI receives greater public trust, improves its reputation and image, and students are more satisfied and more inclined to recommend the institution.

Whatever approach to quality management is adopted, higher education management needs resources and tools to manage ongoing processes effectively. According to Robinson [11], quality management can be supported by information from core functional areas (finance, student records, etc.) and data from monitoring, evaluation, and satisfaction surveys. Researchers argue that quality can be monitored by looking at the impact of higher education in terms of evidence of high-quality student performance, including wages, employer satisfaction, and success in further study. For monitoring purposes, HEI may also collect personal data.

An effective quality assurance system includes continuous quality assessment [10]. While the internal evaluation is the basis for improvement, external evaluation serves as a benchmark, ensures public trust, and conforms to generally accept good practices for organizing and conducting distance learning.

The hardships in defining the meaning of quality pose challenges in developing quality assessment models and tools [50]. Researchers believe that distance learning programs, in addition to specific criteria [51], should be assessed according to the evaluation criteria of full-time programs [52]. Quality can be measured by student engagement and satisfaction [28], [53-54] and their attitude towards distance learning. In addition to evaluating the quality of the training, HEI can use the student feedback to develop and promote courses and programs for teachers' professional development and take results into account during the attestation of the teachers and when making decisions about their promotion.

III. RESULTS: APPROACHES AND TOOLS FOR QUALITY ASSURANCE AND EVALUATION

The known approaches to ensuring and assessing the quality of distance learning can be divided into five main groups:

- Development of normative documents and manuals for quality assurance.
- Conducting audits and accreditation by evaluation organizations.
- Conducting periodic surveys among stakeholders (academic staff, students, external stakeholders).
- Processing of statistical data and performance indicators, including by using intelligent data analysis tools.

In most cases, these approaches are applied mixed to provide a comprehensive set of data allowing the evaluation of the quality of distance learning and taking measures to improve it.

A. Normative Documents and Manuals

To ensure quality, some universities worldwide develop, adopt and implement internal regulatory documents written following current national and regional regulations and laws. An example of such regulatory documents created for the needs of organizing and conducting distance learning in Bulgaria is presented in the book "Quality and Assessment of

e-learning (with good university practices)” [55]. The proposed package of documents includes a Strategy for the development of distance learning, Regulations for organizing and conducting distance learning, a Student Support System, a Handbook on the rights and obligations of students, Methodology for preparation, organization, and conduct of distance learning, Testing, and evaluation system, Guidelines and standards for development and acceptance of learning documentation, Procedures, and regulations for actions in case of complaints, Procedures for punishing and preventing plagiarism attempts, Directory for organizing access to electronic resources, Regulations for technical and technological provision of training, Document for accounting for the specifics of workload, Measures, and procedures for restoring the infrastructure in case of damage or breakdown.

Many independent organizations worldwide promote and support quality improvement, develop and publish quality assurance guidelines and frameworks, disseminate information on good practices for the organization and delivery of distance learning, and encourage the creation of practitioner networks [56] which can contribute significantly to quality assurance. Examples of such organizations are the European Association of Distance Learning Universities (EADTU), the British Open Learning Association, the Canadian Distance Learning Association, the Norwegian Distance Learning Association, and the International Council for Open and Distance Learning (ICDE). The European Commission for Standardization CEN develops frameworks, specifications, and guidelines to improve the quality and transparency of organizations, products, processes, and services for e-learning. There are currently four published documents: CWA 15555:2006 Guidelines and Support for Building application profiles in E-learning; CWA 15660:2007 Providing good practice for E-Learning Quality Approaches; CWA 15661:2007 Providing E-Learning supplies transparency profiles; CWA 16655-1:2013 In LOC – Part 1: Information Model for Learning Outcomes and Competences.

The Higher Learning Commission [57] developed Assessment Guidelines setting out nine hallmarks of distance learning quality and providing suggestions for sample evidence documents for each. The European Association of Distance Learning (EADL) aims to increase the quality of distance learning and provide student benefits. EADL organizes for its members a forum for open discussions on all matters related to distance learning and for the sharing of ideas and good practices. All association members must meet the quality standards and abide by the code. Minimum quality standards include requirements for pre-enrolment, counseling, examinations, face-to-face training, enrollment and contract, management, instruction, and technology-based learning. Across Europe, EADL membership is seen as a mark of quality. The International Council for Open and Distance Education (ICDE) published a comprehensive global review of quality models in online learning [58], which concluded that a systematic quality assurance process is needed to design distance programs. In the strategic plan for the 2021-2024 period, ICDE sets as its goals the advocacy of distance learning worldwide, the promotion of membership in the organization,

promoting quality in digital, open and flexible learning, and ensuring sustainability.

The Australasian Council on Open, Distance, and e-Learning (ACODE) [59] is developing criteria for using technology in higher education. The ACODE proposes 65 performance indicators covering eight thematic areas (Institution-wide policy and governance for technology-enhanced learning; Planning for institution-wide quality improvement of technology-enhanced learning; Information technology systems, services and support for technology-enhanced learning; The application of technology-enhanced learning services; Staff professional development; Staff support for the use of technology-enhanced learning; Student training for the effective use of technology-enhanced learning; Student support for the use of technology-enhanced learning). Each indicator includes scoping statements, good practice statements, a set of performance indicators, and performance measures for each indicator using a five-point rating scale. Each indicator can be used as an independent indicator, or all indicators to be used together for an overall evaluation.

The Asian Association of Open Universities (AAOU, <https://aaou.ouhk.edu.hk/>) develops Quality Assurance Framework for open and distance learning. It contains 107 statements of good practice for achieving quality, divided into ten categories: Policy and Planning; Internal Management; Learners and Learners ‘profiles; Infrastructure, Media and Learning resources; Learner assessment and evaluation; Research and Community Services; Human Resources; Learners Support; Program Design and Curriculum Development; Course Design and Development.

B. Audits and Accreditation by Evaluation Organizations

Sound quality assurance practices combine self-assessment with external quality assessment by quality assurance and assessment organizations. Accreditation by external accreditation ensures that HEI complies with accepted quality standards and can conduct distance education. Accreditation takes place every few years, depending on the accrediting agency. The accreditation process usually includes the following steps:

- Self-assessment.
- An on-site visit by the expert group, which determines the extent to which the HEI fulfills, the accreditation standards based on a review of supporting documents, conducting interviews with staff and students, and observing distance learning activities.
- Development of a written evaluation report describing strengths and recommendations for improvement in terms of accreditation standards.
- Preparing annual reports on the implementation of the recommendations made.

Kirkpatrick believes that accreditation and assessment are valuable for three reasons [38] - it allows governing bodies to identify challenges and take measures to improve curricula, it catalyzes processes to improve the individual capacity and

qualifications of teachers, it gives a sign of quality and excellence in distance learning programs.

The International Organization for Standardization (ISO) develops the ISO/IEC 1976 series of standards, harmonizing the international concept of e-learning quality by describing the processes influencing the achievement and maintenance of e-learning quality [60]. These processes include content and tool creation, service delivery, training and education, monitoring, evaluation, and all life phases from needs analysis to optimization.

The international organization Quality Matters Program (Quality Matters, <https://www.qualitymatters.org/>) develops a series of rubrics that meet the specific needs of different educational sectors. The quality rubric contains eight core standards (Course Overview and Introduction, Learning Objectives, Assessment and Measurement, Instructional Materials, Learning Activities and Learner Interaction, Course Technology, Learner Support, Accessibility, and Usability) and 41 specific standards for evaluating the quality of online and blended courses, explanations on the application of the standards and the relationship between them, a scoring system and a set of tools that facilitate the assessment process. Three certified reviewers review each course and make specific recommendations for course improvement, the implementation of which will contribute to compliance with quality standards. To be certified learning course must receive at least 85% of the possible points. Certified courses receive a stamp with the year of certification valid for 3-5 years.

The European Foundation for Quality in eLearning (EFQUEL, <http://efquel.org/>) seeks to promote good practice and innovation to achieve high-quality learning worldwide. The primary mission of EFQUEL is to increase the quality of e-learning in European countries by providing services and support to all interested parties. According to EFQUEL, the European Quality Assurance System will strengthen the trust in the quality of e-learning and serve as a reference point worldwide. The Foundation believes that classical approaches to quality assessment (such as defining and documenting minimum requirements for infrastructure, staff competence, administrative services, and technical standards are inadequate if the goal of the quality assurance process is to encourage innovation in e-learning. EFQUEL presents a list of elements divided into five parts (Design principles, Agreement to monitor the quality of teaching practice as a whole, General focus on innovation and transformation of the organization and commitment to the competent customer, Principles for conducting negotiations and when it is possible consensus among partners, Agreement on five steps necessary to obtain accreditation), which can serve as a starting point for creating an alternative approach to quality assurance. EFQUEL develops three quality assessment tools – ECBCheck, UNIQUE and Sevaq+.

ECBCheck (ECBCheck, <http://www.ecb-check.org/>) is a certification framework for e-courses and programs developed by EFQUEL. The quality of e-learning courses and programs is assessed in seven areas (information about the organization of the e-learning program, target group orientation, content quality, program/course design, media design, technology,

evaluation and review) with 51 quality criteria, some of which are mandatory. The educational institution is awarded a quality label after evaluating 51 quality criteria divided into four areas (education and training, organizational strategies and innovations, organizational processes, technologies, equipment and infrastructure), some of which the evaluated institution must fulfill. At the first stage of the assessment, the reviewers check the extent to which the institution met all mandatory criteria. The institution passes to the second stage if it fulfills these mandatory criteria. In the following evaluation stage, reviewers assess the optional quality criteria on a four-point scale (0 – not implemented, one – partially implemented, two – adequately implemented, three – excellently implemented). The final result of the evaluation is the percentage ratio of the sum of the optional criteria and the maximum possible points.

UNIQUE [61] is a quality label awarded to a university for the quality use of information and communication technologies. The evaluation focuses on using information and communication technologies to enhance educational provision and learning support throughout the entire breadth of activity of HEIs. The HEIs who apply must meet the standards for program objectives, program structure, content, resources and learning processes. The assessment process takes place in 6 steps – Application, HEI Eligibility check, HEI Self-Assessment, Peer review by a three-person review team, Decision making for Certification by Awarding Body, and Continuous monitoring of ICT policies in line with the recommendations.

The self-assessment model of e-learning quality SEVAQ [61] was developed based on the Kirkpatrick and EFQM models. The model includes a set of criteria and sub-criteria that cover all aspects of the organization. Internal evaluators assess the quality of training and fill out a questionnaire expressing their agreement level with formulated statements. SEVAQ+ is an extension of SEVAQ [66] developed to allow managers and teachers to participate in the self-assessment process in addition to students. The model covers two main aspects: management of the learning process and resources and management of people. The tool offers both a core of questions and opportunities for a personalized assessment. Assessment results are available in real-time and in different formats, from radial graphs that provide a snapshot to raw data that can be imported into other tools. By identifying areas for improvement, the tool enables institutions to track progress from one semester to the next and compare teaching and learning across institutions. Among the main advantages of Sevaq+ is the combination of a robust assessment framework with the flexibility needed to cover a wide range of institutional and individual contexts.

The Open and Distance Learning Quality Council (ODLQC, <http://odlqc.org.uk/>) contributes to ensuring the quality of education, protecting student interests, and developing standards for quality assurance in education. The proposed standards for quality assurance in open and distance learning are divided into six sections: results, resources, support, sales, suppliers, and collaboration.

After analyzing European policies and projects, the good practices of nine national HE evaluation agencies, and studies

in the field, the Swedish National Agency for Higher Education developed the ELQ model [62-63]. The ELQ model contains ten aspects for evaluating the quality of e-learning in higher education - Material/content; Structure/Virtual Environment; Communication, cooperation, and interactivity; Student assessment; Flexibility and adaptability; Support (students and staff); Vision and institutional leadership; Staff qualification and experience; Resources allocation; Holistic and process aspect. For each of these aspects, the authors develop quality criteria in the form of recommendations for taking specific measures to address problems and issues at the institutional level.

The Distance Education Accrediting Commission (DEAC) produces a handbook detailing accreditation standards, policies and procedures. The accreditation handbook sets out expectations for academic quality, educational services, continuous improvement and ethical business practices for institutions proposing distance learning. This handbook contains 57 main components distributed in twelve quality standards covering all aspects and policies of distance education [64]: Institutional Mission; Institution Effectiveness and Strategic Planning; Program Outcomes, Curricula and Materials; Educational and Student Support Services; Student Achievement and Satisfaction; Academic Leadership and Faculty Qualifications; Advertising, Promotional Literature and Recruitment Personnel; Admission Practice and Enrollment Agreements; Financial Disclosures, Cancellations and Refund Policies; Institution Governance; Financial Responsibility; Facilities, Equipment, Supplies, Record Protection and Retention. The accreditation process assesses an institution's ability to meet all accreditation requirements. DEAC expects institutions to provide evidence of compliance with all specified requirements. Feedback on the institution's performance against these standards can help the institution improve the quality of instruction. DEAC approval is seen as a recognition of quality standards.

The European Association of Distance Learning Universities [65] is developing a Quality assessment for e-learning: a benchmarking approach manual [66]. The organization awards an Excellence award to institutions that ensure the high quality of distance learning. The Quality Assurance Agency for Higher Education (QAA) highlighted the importance of considering student workload carefully in module design and ensuring these expectations are consistent and explicit [67].

Governments refer to published quality assurance frameworks as a reference for establishing their national higher education quality assurance systems [68]. From the beginning of the 1990s, they began to promote the adoption of policies and the creation of national or regional quality assurance agencies and to link public funding of education to quality. The assessment, accreditation, and quality control of distance learning in Bulgarian universities and scientific organizations is carried out by the National Agency for Assessment and Accreditation based on relevant criteria systems and procedures for quality assessment and accreditation. The evaluation is going on in two stages (NAOA, 2017) - I. Evaluation of the organization and environment for conducting and maintaining distance learning, II. Evaluation of a concrete

distance learning program. In the self (assessment) report for the organization and the environment for conducting distance education, HEI must provide evidence of compliance with seven criteria related to educational documentation, internal quality system, procedures for developing and updating documentation, methodological standards for documentation, internal normative documents, policy for the development of the scientific and teaching staff, material-technical and information base. In the report on the (self) assessment of a distance learning program, HEI must provide evidence of compliance with three criteria for implementing the procedures for developing and updating study documentation, rules and activities to stimulate student motivation and financial, material-technical and information base.

C. Periodic Surveys among Stakeholders

Research for ensuring the quality of distance learning dates back to decades ago, resulting in several models and approaches proposed for evaluating and assuring the quality of distance learning by various stakeholders.

The E-Learning Maturity Model (eMM) [69] is a capability assessment model for e-learning processes based on the CMMI and SPICE models. Version 2.3 of the model allows the evaluation of 35 e-learning processes divided into five groups giving an idea of the e-learning maturity degree: Learning, Development, Support, Evaluation and Organization. The authors indicate five e-learning maturity levels: Optimization (continuous improvement of e-learning processes in all aspects); Management (ensuring the quality of resources and student outcomes at the exit); Definition (defining eLearning development and maintenance processes); Planning (clear and measurable goals of eLearning projects), Delivery (creating and delivering process outputs).

Ehlers [70] proposes the requirements for the quality of e-learning from learners' perspective structured into seven groups which include 30 dimensions (a set of criteria from the preferences of learners, grouped based on empirical evidence): Tutor Support; Collaboration; Technology; Cost-Expectations-Benefits; Information Transparency of Provider/Course; Course Structure/Presence Courses, and Didactics.

The SEEQUEL quality framework contains an integrated set of criteria for evaluating the quality of e-learning. The framework proposes three main quality criteria (Learning Processes, Learning Resources, and Learning Context) and 137 sub-criteria [70]. Quality criteria that apply to e-learning can be weighted by different users (people or HEI) using a table with two columns. The first column contains a list of criteria (objective dimensions) for determining quality, and in the second column, stakeholders can put a quantitative assessment of the quality criteria, determining its importance for determining the quality of the object (2 – basic criteria, 1 – important criteria, 0 – minor criteria).

The HELEN model [71] allows the evaluation of the quality of the SE based on 46 criteria divided into six dimensions (Supportive Issues, Learner Perspective, Instructor Attitudes, Technical Quality, Information Quality, and Service Quality). The model allows the learning management system to be evaluated only from the student's point of view. Ozkan and

Koseler emphasize the possibility of its expansion to assess the quality of the learning management system also by other stakeholders - developers, administrators, teachers, designers, external experts, etc.

iQTool [72] is a tool for evaluating the quality of teaching in e-learning curricula and the quality of learning materials. The tool allows creating questionnaires, using the questionnaires for quality assessment and has statistical analysis capabilities to improve the quality of learning materials and teaching. The software tool supports four roles – Assessor, QA Manager, Publisher, and Administrator. The repository for evaluation components offers possibilities for storing and retrieving user profiles and evaluated objects, given within the framework of evaluation procedures answers, definition, and calculation of statistical indicators based on the given answers. The build-on repository is based on IMS Digital Repositories Interoperability. HEIs can integrate the tool with learning management systems. The evaluation module acts as an intermediary between the learning management system and the evaluation component and allows retrieving appropriate questionnaires from the repository according to the object type. The tool records the evaluation result in the repository as a document in IMS QTI Results Reporting format with the user's identification numbers, the resource, and the questionnaire used.

Within the framework of the Excellence project, two tools for assessing the quality of e-learning have been developed - QuickScan (for quick orientation) and FullAssessment (for full assessment) [73] with 33 indicators divided into six areas: Strategic management, Curriculum design, Course design, Course delivery, Staff support, Student support. The QuickScan tool provides a quick insight into the strengths of the eLearning delivered and possible areas for improvement. The questionnaire should be completed by a small team, including representatives of different stakeholders: managers, e-course authors, teachers, and students. The team members can also determine how appropriate the indicators are for evaluating quality in the institution. To prove that the answers are based on facts, it is mandatory to accompany them with supporting documents. FullAssessment makes it possible to determine the effectiveness of e-learning programs and the requirements for improvement by having e-learning experts review the supporting documents and, after a site visit, prepare a report on the overall process and recommendations for improvement. Within the Excellence Next project, some of the indicators for evaluating the quality of e-learning have been updated. The number of indicators for quality assessment is 151.

PDPP is an e-course quality assessment model developed based on the CIPP assessment model [74]. The model allows the evaluation of four phases of the e-learning life cycle: Planning (marketing, applicability, target group, course objectives, funding); Development (design, learning materials design, course web page design, flexibility, student interaction, faculty support, technical support, evaluation); Process (technical support, website usage, interaction, evaluation and support during learning, flexibility); Product (student satisfaction, teaching effectiveness, learning effectiveness, and sustainability).

The e-learning self-assessment tool e-Lsa [75] allows evaluation of the quality of organizations offering e-learning through a set of main criteria and sub-criteria covering aspects related to the organization of learning. For each sub-criterion, the authors define a set of measurable indicators. These indicators are formulated as statements, and the quality is measured through self-assessment by various stakeholders. The self-assessment model is divided into two parts. The first part contains 41 indicators for self-evaluation of the course and the learning process by students at the end of the e-course. The second part includes nine indicators for self-evaluation of learning management (by managers and teachers). A corresponding questionnaire has been developed for each of the two parts. At the end of the assessment, the system analyzes the answers and generates a report that allows the manager to identify the strong and weak criteria and identify the reason for the poor results.

The integrated system for evaluation and improvement of the quality of e-learning [33] allows the assessing the quality of the learning management system in a university consortium based on factors divided into five main groups: Learning objects (quality, validity, media, presentation, copyright), Learning Object Design (concept identification, pedagogical style, media enhancements, interactivity, tests and feedback, interaction, content portability standards, content aggregation), Learner Services (identification, portfolio, records for student activity), Program Presentation (design of graphic elements, colour scheme, font, navigation, interface) and Technology Infrastructure (network frequency, end-user system configuration, server configuration, browser, DB connection, technology, operational compatibility). Based on the above five categories, a questionnaire was developed in which the experts had to rate the characteristics and sub-characteristics. When new course content is added, a message is sent to the experts who must evaluate the content, design, and course presentation and suggest changes. In the proposed evaluation framework, student activity (last login, time, course content read during the session, etc.) is monitored, and feedback and suggestions for improvements are sent based on student performance.

The model proposed by Giorgetti, Romero and Vera [76] for evaluating the quality of distance learning is based on the model for accreditation of distance learning programs of the National Commission of University Evaluation and Accreditation for Evaluation and Accreditation of Universities in Argentina CONEAU and Lorenzo Garcia Aretio's integrated distance university evaluation model. The model assesses three dimensions of the conducted distance learning courses: Professional learning (evaluates students' activity during the training), University Management and Administration (measuring the fit between the university's mission, vision, and goals set for continuous improvement), and Student Support (assessing the ability to allocate material and human resources and their management as part of the learning process). The authors suggest that the quality assessment indicators be divided into six main categories (Functionality, Effectiveness, Efficiency, Availability, Information, and Innovation) and arranged in a table. The frequency with which each indicator must be measured is also defined, and a formula is introduced to calculate the indicator value.

Messo [9] offers an approach for assessing the quality of open and distance learning programs from the students' perspective based on qualitative and quantitative methods. Messo proposes to evaluate indicators in seven areas - registration procedures, access to course instructors, administrative processes, course materials, instructional methods, clarity of syllabus, and exam processes. The collected primary data are analyzed using IBM-Statistical Package for Social Sciences (SPSS version 19) by calculating the mean and distribution frequencies and presenting the results in tables, charts and other statistical presentations. The proposed approach was experimented with evaluate the quality of programs at the Open University of Tanzania by 305 students.

Markova, Glazkova, and Zaborova [1] propose a tool for evaluating the distance learning environment from students' perspective and identifying areas where university administrators, teachers, and technicians can improve their work, to ensure high-quality distance learning. Students rate quality indicators in five domains (interaction and collaboration, instructional design and delivery, assessment, student support services, and e-course design) on a five-point Likert scale. With the proposed tool, quality assessment experiments were conducted among 830 students.

Stracke [77] proposes the Open Ed Quality Framework, which conceptualizes the development of quality at three levels (micro, meso, and macro) and in three dimensions (goals, implementations, and achievements). In this framework, learning designers and learning designs are seen as crucial stakeholders and entities that occupy a meso-level role in the implementation dimension and play a significant role in the quality assurance and evaluation process.

Beskrovnyaya, Freidkina, and Vinogradova [78] propose an approach to design tools for monitoring learning outcomes that allow the assessment of shaping competencies most demanded by the labour market. The empirical basis of the study is the results of the analysis of normative and legal documents on distance learning, information published on the Internet about the educational activity of universities using distance learning technologies in the educational process, scientific research on distance learning, materials of personnel selection agencies, means of learning control (Interactive elements in the lecture, Use of materials created based on the theory of the test to test knowledge, Project implementation).

A team from Plovdiv University is developing a range of tools for evaluating the quality of e-courses and digital resources from the perspective of students and experts in distance learning [79]. The questionnaire for students includes 49 questions divided into 11 areas: learning documentation and educational objectives, distance learning provision team, infrastructure, distance learning preparation and delivery, information support, learning materials and activities, communication, assessment, support, design, and recommendations. The questionnaire developed for experts allows them to evaluate the quality of an e-course regarding content (including basic information), positioning (by composition and type), and design (including model, interactivity, multimedia, communicability, performance, ergonomics and functionality in a hardware and software

environment). Each course and digital resource should be evaluated by at least three experts. The questionnaire for experts contains 50 questions from ten areas: learning documentation and educational objectives, distance learning provision team, distance learning implementation infrastructure, training preparation and delivery, learning information support, learning materials and activities, communication, evaluation, support, and design. The questions require a response on a 5-point Likert scale. With the proposed set of tools, 3350 students evaluated the quality of e-learning in 101 e-courses. By providing automated means for synthesis and analysis of the results for all e-courses, the quality of the e-courses has been assessed by professional directions and areas of higher education. Developed software tools for monitoring student activity in conducted surveys and for subsequent analysis of survey results allowing authorized users to generate summary reports, monitor ongoing surveys, and analyze interim data in real-time.

Firdoussi and colleagues [80] conducted a study to evaluate distance learning in Morocco during the COVID-19 pandemic among 3037 students and 231 teachers. The study explores the limitations of e-learning platforms and how public and private universities conduct these activities. Teachers evaluate distance learning in nine areas - Previous Experience with Distance Learning, Distance Learning Platforms, Use of Platforms, Materials Used, Platforms Assessment, Evaluation during the Confinement Period, Distance Learning in the Future the Workload during the Confinement Period, Expectations of E-learning Platforms. Students rate the quality of distance learning in terms of Previous Experience with Distance Education, Internet Connection Quality, Involvement of Teachers, Use of Materials Produced by Professors, Teaching Methods Preferred by Students, Distance Evaluation, Work Timetable, Preferred Type of Education, Resources Used to Better Understand the Course, Devices Used to Follow the Studies from a Distance, Expectations of Distance Education. The survey results are processed using three methods: descriptive analysis, regression analysis, and qualitative response analysis. Microsoft Power BI is used As a data analysis tool to analyze data, visualize it and draw insights.

Lassoued, Alhendawi, and Bashitialshaer [13] conducted a large-scale study to uncover barriers to achieving quality distance learning during the COVID-19 pandemic among 400 professors and 600 students from universities in the Arab world (Algeria, Egypt, Palestine, and Iraq). For this purpose, a questionnaire was developed, which evaluates 14 obstacles in four categories: Personal obstacles, Pedagogical obstacles, Technical obstacles, and Financial and organizational obstacles. The researchers analyze the results to explore the challenges and opportunities to limit them from the perspectives of faculty and students, classify the barriers and identify differences in identified issues to quality in distance learning during the pandemic by faculty and students, and present suggestions for overcoming these obstacles.

As a result of their studies, Jime'nez-Bucarey and his colleagues [42] proposed a model that measures student satisfaction in three dimensions: teachers, technical service, and service. The impact of each dimension on student satisfaction is assessed using a Partial Least Squares Structural

Equation Model (PLS-SEM). An Importance and Performance Map Analysis (IPMA) is performed to identify improvements that need to be done to increase student satisfaction. The model is tested among 1430 students.

Olney, Li and Luo [81] surveyed 220 employees to identify the necessary skills and staff competencies on which HEIs should focus their professional development activities to improve the quality of distance learning. They use a content analysis methodology to analyze the text responses and compare them to the Competency Framework for Instructional Design proposed by the International Board for Teaching, Performance and Learning Standards (IBSTPI). According to the results of the study, the main competencies identified by the participants were designing training interventions, keeping up with design theories, and communicating to manage stakeholders, teams, and projects.

Toubasi et al. [82] developed a tool to assess the quality of distance learning during the COVID pandemic for the needs of universities in Jordan. The questionnaire consists of 58 questions divided into four sections – student demographic characteristics, student attitudes during the distance learning period, student perceptions of distance learning, and quality evaluation using the DELES tool. DELES include 34 indicators that assess the quality of learning in six areas - instructor support, student interaction and collaboration, personal relevance, authentic learning, active learning, and student autonomy. Each indicator is assessed using a 5-point Likert scale. The questionnaire was presented in Google Forms and shared with students through social networks. Results were analyzed with IBM SPSS.

Sarmiento and Callo [53] surveyed to determine the effect of distance learning factors on the quality of learning during the COVID-19 pandemic among 764 students and 57 faculty members. The questionnaire developed contains questions in three parts - profile of respondents, factors and quality learning. Respondents must rate 18 factors that play a crucial role in quality distance learning, divided into three categories (instructional design, support system, implementation), and define purposeful and meaningful distance learning based on engagement, satisfaction, quality teaching, and quality education. The results show that the factors determining the quality of training during the pandemic are instructional design, support system, and implementation.

D. Processing of Statistical Data and Performance Indicators

When conducting distance learning, a lot of data is accumulated about the training, e.g. logs in the e-learning system, data on learning materials read, data on submitted homework assignments, grades from exams, etc. This fact has stimulated research into using data to gain insights into the quality of distance education delivered and support management decisions to retain students and increase student achievement.

Processing such data for the training can provide valuable insights into key performance indicators (e.g. average weekly usage, modules completed, course completion rate, dropout rate, activity completion rate, average attention rate, etc.), highlighting the effectiveness of courses and curricula. By

analyzing retention rates, pass rates, and student satisfaction, institutions can evaluate which learning courses are performing well and which may require improvement [83]. Utilizing big data analytics allows HEIs leadership to do a more sophisticated analysis than simple summary statistics, enabling the identification of courses with pass rates that exceed expectations based on previous student achievements. Through comparative analysis of course designs, HEIs leadership can identify features that lead to successful learning outcomes [84]. They can utilize this knowledge to stimulate teachers to design more efficient courses. In addition, teachers can monitor their learning courses to identify pinch points, such as areas where student engagement drops sharply. They can then take appropriate action, such as providing additional teaching or rewriting course material for the next cohort. Tutors can also use statistical data and performance indicators to monitor their students and identify those who may benefit from timely interventions. By targeting these students, tutors can increase their chances of passing the course and achieving their learning objectives. Students can also use these data to monitor their performance and learning behaviours' [85]. By comparing their progress over time or against other students in their cohort, they can identify areas where they may need to improve and take action as self-regulated learners. Furthermore, teachers can use automated systems to suggest alternative resources or behaviours' to students who exhibit patterns associated with poor results [86]. These suggestions can help students identify areas for improvement and take action to achieve better learning outcomes.

Design, development, and implementation of intelligent data analysis tools can contribute to distance learning quality assurance [87-89]. Automated evaluation of the quality of distance learning requires the collection, analysis, and interpretation of a huge amount of data reflecting the attitude of students and experts to the training courses, the software tools used, etc.

In connection with the external evaluation by accrediting institutions, several studies have been conducted for the automated extraction and analysis of data for distance learning quality evaluation, including data on the used learning materials, infrastructure, e-learning environment, means of communication and collaboration, student evaluation system, flexibility and adaptability of the learning process, student support, team qualification, etc. Doneva and Gaftandzhieva [90] explore the possibilities for automated data extraction for evaluating the criteria from the criteria system of NAOA for evaluating distance learning programmes were analyzed. This analysis aims to determine which data can be extracted from university systems (e-learning environment, learning process management system, academic staff development system, etc.) to support distance learning quality assessment. Based on the analysis, some experimental web services are developed for extracting data from the Moodle e-learning environment for automated evaluation of the quality of distance learning. The following work [79] proposes the automation of related processes based on an approach for integrating heterogeneous software systems (Service Oriented Integration) and discusses its application to automated data extraction in evaluating the quality of e-learning. Some reports with extracted data from

the learning management system are generated through the developed tools and presented to an expert group during program accreditation of a distance learning program in 2016. Among them are the reports for Activity in communication tools, Study schedule control, Student and teacher workload, Student success, and Educational activities and resources.

As a result of studies in the field and after analysis of the databases of the university information systems, in which data about the training are stored, Gaftandzhieva and Doneva [91] propose a model with a set of indicators, allowing the tracking of training results for the needs of various interested parties (students, teachers, program managers, faculty leadership, university leadership, and quality experts). Based on the indicators from the proposed model, four tools for intelligent data analysis to improve learning outcomes have been designed and developed. The mobile application Mobile LAP [92] allows students to track the values of indicators (for student activity, control of the study schedule, and student success) that can help them achieve their goals in study time and improve their success. As use Mobile LAP, students can track their activity and progress and compare it to the results of other students, as well as monitor whether they are following the study schedule. LATeach application [93] allows teachers to track student activity in learning activities, compare the results of a selected student with those of other students in the course and with the results of students who received excellent grades in previous years, monitor student compliance with the study schedule, track student progress in learning activities and learning outcomes and student success during the learning process, identify at-risk students and self-assess the quality of learning resources based on the students' activity and their results. LATch tool [94] allows the governing bodies in HEIs to generate reports with aggregated data on the students' activity and success rate in selected study programs, which allows them to track the results of students and compare them with those of students from previous years, to identify programs in which students are not performing satisfactorily, to track trends in student success by comparing students' GPA at the end of each academic year, to track student success at graduation, to track student dropout rates, etc. The generated reports for each indicator enable university management at different levels to make informed decisions to improve the quality of education and the results achieved. The LAqe tool [95] allows quality experts to generate dynamic reports for monitoring and evaluating the quality of conducted education for the needs of accreditation procedures. They can use the tool to generate evidence documents and significantly support the preparation of self-assessment reports for internal and external evaluation of the quality of the training provided.

IV. CONCLUSION

This review paper provided a comprehensive analysis of the current state of play of quality assurance in distance learning. The paper has highlighted the challenges associated with ensuring quality in distance learning programs and has discussed various approaches and tools for quality assurance and assessment, including regulatory documents and manuals, audits, and stakeholder satisfaction. Using statistical data and monitoring key performance indicators, HEIs can identify areas for improvement and take appropriate action to enhance

the quality of their distance learning programs. As distance learning continues to grow in popularity, institutions must prioritize quality assurance to ensure that their programs meet established standards and provide students with a high-quality education. Subsequently, this paper serves as a valuable resource for educators, administrators, and policymakers interested in improving the quality of distance learning programs. By implementing the recommendations outlined in this paper, HEIs can enhance the quality of their distance learning programs and provide students with a more rewarding and fulfilling learning experience.

This study was conducted as part of a project to implement software tools to help ensure the quality of educational and administrative services at the university and support management decision-making to ensure high quality of services. In the next part of the research, tools will be designed and developed to track the values of key performance indicators for the needs of different stakeholders (teachers, distance learning centres, dean's and rectors' leaderships).

ACKNOWLEDGMENT

This paper is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01. The paper reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

- [1] T. Markova, I. Glazkova, E. Zaborova, "Quality Issues of Online Distance Learning", 7th Procedia - Social and Behavioral Sciences, 237, 685 – 691, 2017.
- [2] J. Bennett, L. Bennett, "Assessing the quality of distance education programs: The faculty's perspective", Journal of Computing in Higher Education, 13, 71-86, 2022
- [3] A. Nsamba, M. Makoe, "Evaluating quality of students' support services in open distance learning", Turkish Online Journal of Distance Education, 18(4), 91- 103, 2017.
- [4] K. Greenan, "Student Engagement in the Virtual Classroom: Implications for Overcoming Conflict Between Instructors and Students and Creating Collaborative Virtual Workspaces". Contemporary Trends in Conflict and Communication: Technology and Social Media, 209, 2022.
- [5] B. Bokayev, Z. Torebekova, Z. Davletbayeva, F. Zhakypova, "Distance learning in Kazakhstan: estimating parents' satisfaction of educational quality during the coronavirus", Technology, Pedagogy and Education, 30(1), 27-39, 2021.
- [6] L. Car, B. Kyaw, R. Panday, R. van der Kleij, N. Chavannes, A. Majeed, J. Car, "Digital health training programs for medical students: scoping review". JMIR Medical Education, 7(3), e28275, 2021.
- [7] I. Aydin Su'nbu, M. Aslan Go'rdesli, "Psychological capital and job satisfaction in public-school teachers: the mediating role of prosocial behaviours", Journal of Education for Teaching, 47(2), 147-162, 2021.
- [8] T. Alodwan, "Online Learning during the COVID- 19 Pandemic from the Perspectives of English as Foreign Language Students", Educational Research and Reviews, 16(7), 279-288, 2021
- [9] I. Messo, "Students' perception on the quality of open and distance learning programmes in Tanzania", Huria: Journal of the Open University of Tanzania, 18(1), 119-134, 2014.
- [10] A. Zuhairi, M. Raymundo, K. Mir, "Implementing quality assurance system for open and distance learning in three Asian open universities: Philippines, Indonesia and Pakistan", Asian Association of Open Universities Journal, 15(3), 297-320, 2020.

- [11] B. Robinson, "The Management of Quality in Open and Distance Learning.", Proceedings of the Eighth Annual Conference of the Asian Association of Open Universities, New Delhi, February 20-22, 1995. Vol. 1, 95-109, 1995
- [12] M. Burns, "Distance education for teacher training: Modes, models, and methods". Education Development Center. Inc. Washington, DC, 338, 2011.
- [13] Z. Lassoued, M. Alhendawi, R. Bashitialshaaer, "An exploratory study of the obstacles for achieving quality in distance learning during the COVID-19 pandemic", Education sciences, 10(9), 232, 2020.
- [14] R. Roe, "Considering quality control in distance and online education: A commentary", Kentucky Journal of Excellence in College Teaching and Learning, 8(1), 7, 2011.
- [15] F. Henard, D. Roseveare, "Fostering Quality Teaching in Higher Education: Policies and Practices an IMHE Guide for Higher Education Institutions, OECD, Paris, 2012.
- [16] M. Thorpe, "Rethinking learner support: The challenge of collaborative online learning", Open learning, 17(2), 105-119, 2002.
- [17] D. Valentine, "Distance learning: Promises, problems and possibilities", Online Journal of Distance Learning Administration, 5(3), 2002
- [18] K. Elumalai, J. Sankar, J. John, N. Menon, M. Alqahtani, M. Abumelha, "Factors Affecting the quality of e-learning during the COVID-19 Pandemic from the perspective of higher education students", Journal of Information Technology Education: Research, 19, 731-753, 2020
- [19] D. Vlachopoulos, M. Agoritsa, "Quality Teaching in Online Higher Education: The Perspectives of 250 Online Tutors on Technology and Pedagogy. International Journal of Emerging Technologies in Learning, 16(06), 40-56, 2021.
- [20] J. Ann-Kathrin, K. Leibniz, R. Gollner, "Distance Teaching During the COVID-19 Crisis: Social Connectedness Matters Most for Teaching Quality and Students' Learning", AERA Open, 7(1), 1-14, 2021.
- [21] S. Affouneh, S. Salha, Z. Khlaif, "Designing quality e-learning environments for emergency remote teaching in coronavirus crisis", Interdiscip J Virtual Learn Med Sci, 11(2), 1-3, 2020.
- [22] T. Favale, F. Soro, M. Trevisan, I. Drago, M. Mellia, "Campus traffic and e-Learning during COVID-19 pandemic", Computer Networks, 176(April), 2020, <https://doi.org/10.1016/j.comnet.2020.107290>
- [23] M. Sadeghi, "A shift from classroom to distance learning: Advantages and limitation", International Journal of Research in English Education (IJREE), 4(1), 80-88, 2019.
- [24] G. Lorenzo, J. Moore, "The Sloan Consortium Report to the Nation: Five Pillars of Quality Online Education", 2002.
- [25] M. Allen, E. Mabry, M. Mattrey, J. Bourhis, S. Titsworth, N. Burrell, "Evaluating the Effectiveness of Distance Learning: A Comparison Using Meta-Analysis", Journal of Communication, 54(3), 402-420, 2004.
- [26] A. Ni, "Comparing the effectiveness of classroom and online learning: Teaching research methods", Journal of Public Affairs Education, 19(2), 199-215, 2013.
- [27] T. Elias, "Universal instructional design principles for Moodle". International Review of Research in Open and Distance Learning 11(2), 2010.
- [28] J. McClary, "Factors in high quality distance learning courses", Online Journal of Distance Learning Administration, 16(2), 230-256, 2013
- [29] J. Lee, C. Dziuban, "Using quality assurance strategies for online programs. Educational Technology Review, 10(2), 69-78, 2002.
- [30] I. Grutzner, S. Weibelzahl, P. Waterson, "Improving Courseware Quality through Life-Cycle Encompassing Quality Assurance, Proc. of Symposium on Applied Computing (SAC'04), 946-951, 2004.
- [31] C. Wright, "Criteria for Evaluating the Quality of Online Courses", Instructional Media and Design, Grant MacEwan College, Edmonton, Alberta, Canada, 2004.
- [32] L. Halverson, C. Graham, "Learner engagement in blended learning environments: A conceptual framework", Online Learning, 23(2), 145-178, 2019.
- [33] L. She, L. Ma, A. Jan, N. Sharif, P. Rahmatpour, "Online Learning Satisfaction During COVID-19 Pandemic Among Chinese University Students: The Serial Mediation Model", Front. Psychol. 12:743936, 2021.
- [34] W. Zaman, P. Ghosh, K. Datta, P. Basu, "A Framework to Incorporate Quality Aspects for e-Learning System in a Consortium Environment", International Journal of Information and Education Technology, 2(2), 2012..
- [35] C. Montedoro, C., V. Infante, "Linee Guida per la Valutazione di Qualita' del Software didattico nell'e-Learning: ISFOL", I libri del Fondo Sociale Europeo, Roma, 2003.
- [36] R. Lanzilotti, C. Ardito, M. Costabile, A. De Angeli, "eLSE Methodology: a Systematic Approach to the e-Learning Systems Evaluation", Educational Technology Society, 9 (4), 42-53, 2006.
- [37] T. Markova, I. Glazkova, E. Zaborova, "Quality issues of online distance learning". Procedia-Social and Behavioral Sciences, 237, 685-691, 2017.
- [38] D. Kirkpatrick, "Quality assurance in open and distance learning", 2005.
- [39] I. Kusmaryono, W. Kusumaningsih, "A Systematic Literature Review on the Effectiveness of Distance Learning: Problems, Opportunities, Challenges, and Predictions", International Journal of Education, 14(1), 62-69, 2021
- [40] F. Ferraro, F. Ambra, L. Aruta, M. Iavarone, "Distance learning in the covid-19 era: Perceptions in Southern Italy". Education Sciences, 10(12), 1-10, 2020.
- [41] P. Fidalgo, J. Thormann, O. Kulyk, J. Lencastre, "Students' perceptions on distance education: A multinational study", International Journal of Educational Technology in Higher Education, 17(1), 1-15, 2020.
- [42] C. Jimenez-Bucarey, A. Acevedo-Duque, S. Muller-Perez, L. Aguilar-Gallardo, M. Mora-Moscoso, E. Vargas, "Student's Satisfaction of the Quality of Online Learning in Higher Education: An Empirical Study", Sustainability, 13, 11960, 2021.
- [43] Y. Klepalova, I. Yur, Y. Tarasova, O. Savka, E. Maistrovich, "Reliability and quality of distance learning technical education in the context of the COVID-19 pandemic: practice and issues", Journal of Physics: Conference Series, 2001(1), 012036, 2021.
- [44] G. Lethuillier, P. Nkengne, "The challenge of monitoring quality in distance education", IIEP-UNESCO Dakar, 2020.
- [45] G. Alfonso, "Quality assurance in distance education", Papers of UPOU Chancellors, Open University, Los Banos, 126-130, 2015.
- [46] T. Belawati, A. Zuhairi, "The practice of a quality assurance system in open and distance learning: a case study at Universitas Terbuka Indonesia (The Indonesia Open University)", International Review of Research in Open and Distance Learning, 8(1), 1-15, 2007
- [47] T. Bibi, I. Rokhiyah, D. Mutiara, "Comparative study of quality assurance practices in open distance learning (ODL) universities", International Journal of Distance Education and E-Learning, 4(1), 26-39, 2018.
- [48] N. Jamandre, "Quality assurance in distance education achieved in the Philippines", Asian Journal of Distance Education, 9(1), 90-97, 2011.
- [49] I. Jung, T. Wong, T. Belawati, "Quality Assurance in Distance Education and E-Learning Challenges and Solutions from Asia", Sage Publications, New Delhi and IDRC, Ottawa, 2013.
- [50] L. Harvey, D. Green, "Defining quality", Assessment and Evaluation in Higher Education, 18(1), 9-34, 1993.
- [51] A. Stella, A. Gnanam, "Quality assurance in distance education: the challenges to be addressed", Higher Education, Vol.47, 143-160, 2004
- [52] P. Whiteley, "Assessing the quality of distance education: the case of the university of the West Indies.
- [53] M. Sarmiento, E. Callo, "Learning Quality of Senior High School Distance Education During the COVID-19 Pandemic". International Journal of Educational Management and Development Studies, 3(4), 2022.
- [54] A. Valai, D. Schmidt-Crawford, K. Moore, "Quality Indicators for Distance Learning: A Literature Review in Learners' Perceptions", Proceedings of International Journal on E-Learning, 103-124, 2019.
- [55] S. Uvalic-Trumbic, S. Daniel, "A Guide to Quality in Online Learning", Academic Partnerships, Mountainview, California, 2013.

- [56] E. Huertas, I. Biscan, C. Ejsing, L. Kerber, L. Kozłowska, S. Ortega, L. Lauri, M. Risse, K. SchCorg, G. Seppmann, "Considerations for Quality Assurance of Learning Provision Report from the ENQA Working Group VIII on Quality Assurance and Learning Occasional Papers 26, European Association for QA in Higher Education AISBL, Brussels, Belgium, 2018.
- [57] Higher Learning Commission, "Guidelines for the evaluation of distance education (on-line learning)", Chicago: HLC, 2009.
- [58] E. Ossiannilsson, K. Williams, A. Camilleri, M. Brown, "Quality Models in Online and Open Education Around the Globe State of the Art and Recommendations", International Council for Open and Distance Education-ICDE, Oslo, 2018.
- [59] ACODE, "Benchmarks for Technology Enhanced Learning", 2014.
- [60] ISO/IEC 19796-1:2005 Information technology – Learning, education and training – Quality management, assurance and metrics – Part 1: General approach.
- [61] J. Schreurs, A. Husson, B. Merison, E. Morin, H. Van Heysbroeck, "SEVAQ: A Unique Multi-functional Tool for Assessing and Improving the Quality of e-Courses", International Journal of Emerging Technologies in Learning, 3(1), p.61, 2008.
- [62] U. Ehlers, C. Helmstedt, M. Bijmens, "Shared Evaluation of Quality in Technology-enhanced Learning", Whitepaper, 2011.
- [63] SNAHE, "E-learning quality: Aspects and criteria for evaluation of e-learning in higher education", Swedish National Agency for Higher Education, 2008
- [64] H. Hansson, P. Westman, E. Astrom, M. Johansson, "Aspects and criteria for evaluation of e-learning in higher education", Swedish National Agency for higher Education E-learning quality Report, 11, 2008.
- [65] Distance Education Accrediting Commission, "DEAC Accreditation Handbook", 2022.
- [66] EADTU, "Quality Assessment for E-Learning: A Benchmarking Approach", 3rd ed., European Association of Distance Teaching Universities, Maastricht, 2016.
- [67] QAA, "Explaining contact hours", 2011.
- [68] L. Brockerhoff, J. Huisman, M. Laufer, "Quality in Higher Education: A Literature Review", Centre for Higher Education Governance, Ghent University, Ghent, Belgium, 2015.
- [69] S. Marshall, G. Mitchell, "Benchmarking International E-learning Capability with the E-Learning Maturity Model", Proceedings of EDUCAUSE in Australia, 2007.
- [70] U. Ehlers, I. Pawłowski, "Handbook on quality and standardization in e-learning", 2006.
- [71] S. Ozkan, R. Koseler, "Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation", Computers & Education, 53, 1285–1296, 2009
- [72] N. Moumoutzis, M. Christoulakis, P. Arapi, M. Mylonakis, S. Christodoulakis, "The iQTool Project: Developing a Quality Assurance Tool for Elearning", LOGOS Open Conference on strengthening the integration of ICT research, 202-210, 2009.
- [73] K. Williams, K. Kear, J. Rosewell, Quality Assessment for E-learning: a Benchmarking Approach, Second Edition, 2012.
- [74] W. Zhang, Y. Cheng, "Quality Assurance in E-Learning: PDPP Evaluation Model and its Application", The International Review of Research in Open and Distance Learning, Research Articles, 13(3), 66-82, 2012.
- [75] J. Schreurs, A. Al-Huneidi, "An eLearning Self-Assessment Model (e-LSA)", Proceedings of the fifth international conference on e-learning in the workplace, New York, USA, 2012.
- [76] C. Giorgetti, L. Romero, M. Vera, "Design of a specific quality assessment model for distance education", Universities and Knowledge Society Journal, 10(2), 301-315, 2013.
- [77] C. Stracke, "Quality frameworks and learning design for open education", International Review of Research in Open and Distributed Learning, 20(2), 180-202, 2019.
- [78] V. Beskrovnaya, E. Freidkina, T. Vinogradova, "Approaches to assessing the quality of distance learning in higher education through the development of tools for monitoring learning outcomes", International scientific and practical conference on digital economy, 433-438, 2019.
- [79] S. Gaftandzhieva, G. Totkov, R. Doneva, "Quality and evaluation of e-learning (with good university practices)", Paisii Hilendarski University Publishing House, 424 pages, 2020.
- [80] S. El Firdoussi, M. Lachgar, H. Kabaili, A. Rochdi, D. Goujdami, L. El Firdoussi, "Assessing distance learning in higher education during the COVID-19 pandemic". Education Research International, 2020, 1-13.
- [81] T. Olney, C. Li, J. Luo, "Enhancing the quality of open and distance learning in China through the identification and development of learning design skills and competencies". Asian Association of Open Universities Journal, 2021.
- [82] A. Toubasi, S. Al-Harasis, Y. Obaid, F. Albustanji, H. Kalbouneh, "Quality of Distance Learning After One and a Half Year From Its Integration Due to the COVID-19 Pandemic: A Cross-Sectional Study at the University of Jordan". Cureus, 14(12), 2022.
- [83] M. Yanez Pagans, "Disclosable Restructuring and (or) Additional Financing Paper-Strengthening the Capacity to Produce and Use Quality Education Statistics- P163049", 2020.
- [84] M. Saini, E. Sengupta, M. Singh, H. Singh, J. Singh, "Sustainable Development Goal for Quality Education (SDG 4): A study on SDG 4 to extract the pattern of association among the indicators of SDG 4 employing a genetic algorithm". Education and Information Technologies, 28(2), 2031-2069, 2023.
- [85] E. Ivanova, I. Vinogradova, S. Zadadaeva, "The study of school educational environment in the context of ensuring equal access to quality education", The Education and science journal, 21(7), 69-89, 2019.
- [86] W. Tang, X. Zhang, Y. Tian, "Mitigation of Regional Disparities in Quality Education for Maintaining Sustainable Development at Local Study Centres: Diagnosis and Remedies for Open Universities in China", Sustainability, 14(22), 14834, 2022.
- [87] L. Lockyer, E. Heathcote, S. Dawson, "Informing pedagogical action: aligning learning analytics with learning design", American Behavioural Scientist, 57(10), 2013.
- [88] R. Galley, "Learning design at the Open University: introducing methods for enhancing curriculum innovation and quality", Quality Enhancement Report Series, Vol. 1, 2015.
- [89] J. Dalziel, G. Conole, S. Wills, S. Walker, S. Bennett, E. Dobozy, L. Cameron, B. Badilescu, M. Bower, "The Larnaca declaration on learning design", Journal of Interactive Media in Education, 1(7), 1-24, 2016.
- [90] R. Doneva, S. Gaftandzhieva, "Automated e-learning quality evaluation", eLearning'15 Proceedings of the International Conference on e-Learning, 156-162, 2015.
- [91] S. Gaftandzhieva, R. Doneva, "A Comprehensive Approach to Learning Analytics", ICERI2019 Proceedings, 2634-2643, 2019.
- [92] S. Gaftandzhieva, R. Doneva, S. Petrov, G. Totkov, "Mobile Learning Analytics Application: Using Students' Big Data to Improve Student Success", International Journal on Information Technologies & Security, 10(3), 53-64, 2018.
- [93] S. Gaftandzhieva, R. Doneva, G. Pashev, "Learning Analytics from the teacher's perspective: a mobile app", INTED2019 Proceedings, 8133-8143, 2019.
- [94] R. Doneva, S. Gaftandzhieva, M. Bliznakov, S. Bandeva, "Learning Analytics Software Tool Supporting Decision Making in Higher Education", International Journal on Information Technologies and Security, 12(2), 37-46, 2020.
- [95] S. Gaftandzhieva, R. Doneva, M. Bliznakov, "Internal and External QA in HE: LA Tools and Self-Evaluation Report Preparation", International Journal of Emerging Technologies in Learning, 15(16), 191-199, 2020.

Efficient Parameter Estimation in Image Processing using a Multi-Agent Hysteretic Q-Learning Approach

Issam QAFFOU

ISI Laboratory-Department of Computer Science-Faculty of Sciences Semlalia, Cadi Ayyad University
Boulevard Prince My Abdellah B.P. 2390 | 40000 Marrakech. Morocco

Abstract—Optimizing image processing parameters is often a time-consuming and unreliable task that requires manual adjustments. In this paper, we present a novel approach that utilizes a multi-agent system with Hysteretic Q-learning to automatically optimize these parameters, providing a more efficient solution. We conducted an empirical study that focused on extracting objects of interest from textural images to validate our approach. Experimental results demonstrate that our multi-agent approach outperforms the traditional single-agent approach by quickly finding optimal parameter values and producing satisfactory results. Our approach's key innovation is the ability to enable agents to cooperate and optimize their behavior for the given task through the use of a multi-agent system. This feature distinguishes our approach from previous work that only used a single agent. By incorporating reinforcement learning techniques in a multi-agent context, our approach provides a scalable and effective solution to parameter optimization in image processing.

Keywords—Parameter estimation; reinforcement learning; cooperative agents; hysteretic q-learning; optimistic agent; object extraction

I. INTRODUCTION

Image processing tasks often require the application of one or more image processing operators that are parameterized, requiring assignment of values to the parameters. However, changing parameter values can significantly impact the quality of the processing result. Non-expert users may face challenges in manually computing optimal parameter values, particularly when multiple operators are involved. For instance, the Deriche filter [1] is frequently used to detect contours in an image, and its parameter α represents the size of the filter. Non-expert users may need to make several attempts to find satisfactory results, wasting time and computing resources.

To address this issue, we propose a novel approach to automatically estimate parameters in image processing using a multi-agent system and reinforcement learning. Our approach leverages cooperative learning between agents to outperform centralized learning. The main contribution of this paper is the use of a multi-agent system to tackle the challenge of parameter estimation in image processing.

In this paper, we begin by discussing related works in Section II. Section III introduces preliminaries, while Section IV details the proposed approach. In Section V, we present experiments and results to validate our approach's effectiveness. Finally, in Section VI, we conclude the paper and discuss potential future work.

II. RELATED WORKS

The problem of parameter estimation in image processing has garnered interest from various researchers. In their work [2], Elie Zemmour et al. proposed an automatic method for estimating the parameters required for adaptive thresholding to detect peppers and apples in varying lighting conditions. The authors focused on the adjustment of light level threshold, stop splitting conditions, and classification rule direction for detecting the specific objects (apples and peppers) under consideration. However, the proposed algorithm's adaptability to other image processing tasks was not discussed and is likely to depend on the specific object characteristics. Rafael et al. [3] introduced a method that uses a racing algorithm to tune the parameters required for document image binarization. Their approach is based on a statistical method for determining the optimal parameter values for two algorithms used in binarization tasks: the perception of objects by distance and its combination with a Laplacian energy-based method. Meanwhile, in a study aimed at improving image and video codecs that widely employ uniform quantization schemas, Miguel et al. [4] identified the size of the dead zone and the reconstruction point location as the critical parameters affecting the image R/D coding. The authors proposed a parameterized Uniform Variable Dead Zone Quantizer (UVDZQ) for encoding using wavelets, and evaluated its performance against the most popular quantizers used in video and image coding - USQ (Uniform Scalar Quantizer) and USDZQ (Uniform Scalar Dead Zone Quantizer). The optimal display of an image requires an optimal gamma transformation. In [5], Wang et al. proposed a method based on the location of the Zero-Value Histogram Bin (ZVBH) to estimate the gamma transformation parameter. This approach leverages the relationship between the parameter and the number of ZVBHs to approximate the optimal parameter value and its associated interval. When it comes to biological applications, the choice of parameter values can impact the results obtained. Diana B. et al. [6] proposed using Gaussian process learning to estimate the best biological parameters from non-quantitative and noisy image data. They validated their approach on a parametric function and applied it to estimate parameters in a biological setting by adjusting artificial ISH (in-situ hybridization) data of the developing murine limb bud. Machine learning has become a popular solution to optimization problems in image processing, including parameter estimation. In their work [7], Qaffou et al. proposed an automatic solution for adjusting parameters in an object recognition task using the Q-learning algorithm [54]. This algorithm allows the agent to identify the optimal

combination of parameters for two vision operators - GLCM (Gray Level Co-occurrence Matrix) and k-means. In another study [8], the authors proposed a general framework for optimizing the process of operator and parameter estimation independently on a specific image processing task using reinforcement learning. Furthermore, in [9], Qaffou et al. explored a multi-agent architecture for modeling the interaction between the three components of the proposed architecture. Their solution was successfully applied to a segmentation task, and the results demonstrated its ability to adapt to user preferences [10]. However, the multi-agent architecture proposed in [9, 10] only models different types of agents but has only one agent that learns the optimal values. Qingang et al. [11] proposed a decoupled learning methodology that dynamically fits the weights of a deep network as most existing trained models rely on the configuration of a single parameter. Jinming et al. [12] proposed a simple method for learning local parameter tuning in adaptive image processing by extracting local characteristics from an image and learning the relationship between them and the optimal filtering parameters, optimizing any metric that defines the image's quality.

III. PRELIMINARIES

Most image processing tasks involve the use of one or more vision operators, which are typically parameterized. Each parameter has a range of possible values, and in order to execute an operator, the user must assign a value to each parameter. If the resulting output is unsatisfactory, the user may try other parameter values or even switch to a different operator altogether. In this paper, we focus on the parameter estimation process assuming that the operators to be used have already been fixed. The approach proposed in this paper requires the introduction of certain concepts, which are explained in the following subsections.

A. Overview

We assume that image processing tasks require the use of multiple operators, each of which is parameterized with a range of values. The processing is guided by the ground truths provided for the input images. Our proposed approach in this paper is based on the concepts of multi-agent systems, and requires the following components for the agents to function effectively:

- The combination of operators to use.
- The range of possible values for each parameter.
- The input images.
- The image reference which serves as the ground truth for the desired result. For example, in the case of segmentation, the ground truth is a manual segmentation done by an expert, and it is used for evaluation purposes.

Fig. 1 provides a summary of the inputs required for the

multi-agent system used, as well as the output it generates.

B. Multi-agent System

An agent is any autonomous entity that interacts with its environment through sensors and actuators [15]. When such an agent tries to optimize its performance measure, it is called a rational agent. Although intelligent agents are autonomous, they may sometimes need to collaborate and cooperate to complete tasks that require integration or are time-consuming. Agents may or may not cooperate, share knowledge with each other, depending on the task at hand and users' preferences. A system composed of a group of agents capable of interacting with each other is called a multi-agent system (MAS), which constitutes the core of distributed artificial intelligence (DAI) that emerged in the 1980s [13, 14]. Since then, researchers have used MAS to solve problems of distributed or parallel processing in image processing [16-21].

In this study, the multi-agent system used is composed of a team of agents that cooperate using reinforcement learning to speed up the process of finding the optimal values for parameters in image processing tasks. The integration of MAS and reinforcement learning to solve such an optimization problem is an innovative idea.

1) *Reinforcement Learning (RL)*: RL is the technique that forms the basis of the solution proposed in this paper to solve the problem of parameter estimation for a combination of vision operators. We chose this technique because of its adaptability to the dynamicity of environments due to the balance it allows between exploring the environment and exploiting possible solutions [22]. RL was originally developed for Markov Decision Processes (MDPs). RL defines a type of interaction between an agent and its environment, as shown in Fig. 2. In a real state s of the environment, the agent chooses and performs an action a which causes a transition to the state s' . The agent receives a reinforcement signal "r" that is a reward if the action is beneficial or a punishment if not; a null signal means an inability to award a penalty or reward. The agent then uses this signal to improve its strategy, which is the sequence of its actions, in order to maximize the accumulation of future rewards.

To achieve this, the agent must find a balance between exploration and exploitation. Exploration consists of testing new actions that can lead to higher gains, but with the risk that they will be lower, while exploitation consists of applying the best strategy acquired until then (which may not be optimal). Fig. 2 shows the general architecture of an RL agent. There are several methods to find the optimal policy corresponding to the maximum value of the state/action value function. In this paper, the proposed MAS exploits the idea of the Q-learning algorithm, where the agents cooperate and update their Q-values optimistically.

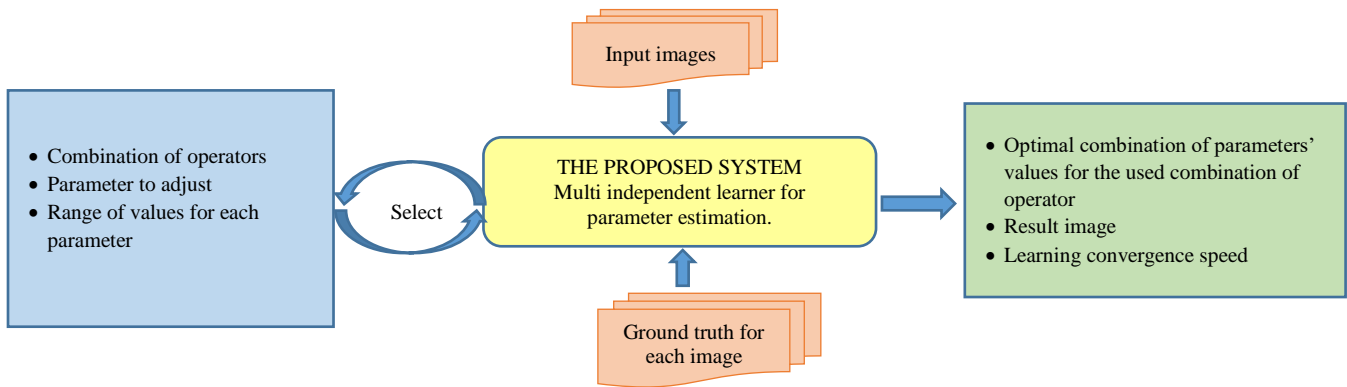


Fig. 1. External architecture of the proposed solution.

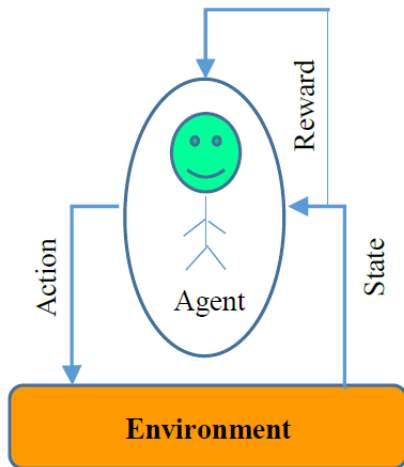


Fig. 2. RL agent general architecture.

2) *Q-learning algorithm*: The Q-learning algorithm was proposed by Watkins in 1989 [23]. In this model, the agent learns to act optimally in Markov domains by testing action sequences. It selects an action in a particular state and uses the immediate reward or punishment to estimate the value of that state. By trying different actions in different states, the agent learns which is the best by referring to the long-term update of the rewards [24]. The agent must determine an optimal policy and maximize the total expected rewards. Algorithm 1 gives the procedural form of the Q-learning algorithm [56].

Alg. 1: Q-learning algorithm

1. Initialize $Q(s, a)$ for all $s \in S, a \in A(s)$ randomly
2. Observe the initial state s
3. Repeat until termination:
 4. Select an action a using a policy derived from Q
 5. Take action a , observe reward r and new state s'
 6. Update

$$Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$
 7. $s \leftarrow s'$
8. Until s is a terminal state

In this algorithm, S is the set of possible states, $A(s)$ is the set of possible actions in state s , α is the learning rate, and γ is the discount factor used to balance immediate and future

rewards. The value of $Q(s, a)$ is updated based on the reward received and the estimated value of the next state-action pair. The policy derived from Q is used to select the action in step 4.

3) *Cooperative learning*: G Although reinforcement learning (RL) has promising applications in multi-agent systems (MAS), there are still several challenges to extending RL to a MAS [25]. One major difficulty is the lack of theoretical guarantees, as convergence hypotheses that hold for a single agent may not be valid for a MAS due to the presence of several learners, which makes the environment non-stationary and challenging for multi-agent learning systems [26-29]. Additionally, defining a good learning goal for all RL agents and enabling communication between them pose further challenges for learning in a MAS [30]. Despite these obstacles, researchers have successfully integrated RL into MAS and addressed coordination problems between agents [31-36]. One algorithm that addresses the challenge of cooperation between agents is hysteretic Q-learning, which focuses on optimistic agent behavior and has been shown to perform well in multi-agent environments [29]. We use this algorithm to include RL in our cooperative MAS. The next section provides more details on how we model our cooperative MAS.

IV. THE PROPOSED APPROACH MODELING

To accomplish a task in image processing, such as filtering, image enhancement, segmentation, object recognition, etc., a sequence of operators must be applied. A task could be divided into several sub-tasks that may require a sub-sequence of vision operators to be used. To run an operator, its parameters must be adequately tuned. This problem has been solved using a single agent [7], but it consumes much time to converge. The main contribution of this paper is to propose a quicker and more precise solution to this problem. We model our solution as a multi-agent system using reinforcement learning. Each agent takes one operator and learns to find the optimal values of its parameters depending on the user's preferences. These preferences include the type of process the user wants to perform, the operators they want to use, and the desired result. The agents work together to find the best choice that brings them the highest reward. These agents have a joint action, and each one has its

individual action in the whole process. An agent's individual actions are formed by all possible choices of the values that can be assigned to its parameters. Any value changed is a new action. Formally speaking, if an operator has, for instance, k parameters (p_1, p_2, \dots, p_k) and each parameter p_i has a set of possible values $V_i = \{v_{i_1}, v_{i_2}, \dots, v_{i_p}\}$, an individual action is then:

$$a_j = (u_1, u_2, \dots, u_k) \in V_1 \times V_1 \times \dots \times V_k$$

Each agent chooses its action independently of the choices of the other agents. For the multi-agent system, we talk about a joint action which is a combination of the individual actions. It is the action that the multi-agent system applies on the input image, which represents the environment with which the agents are interacting. A state of the environment is a set of features extracted from the image. For each image, we provide a ground truth to evaluate the result found by the proposed multi-agent system. The reward is calculated by comparing the features of the output image with those of its ground truth. These components, which are the set of actions, states, and the reward function, are the principal elements required to define the reinforcement learning process. Fig. 3 shows the global schema of the proposed multi-agent system. The proposed multi-agent system applies a joint action on the input image, compares the obtained result with the ground truth, and receives a return signal in the form of a reward or punishment. During learning, the system reinforces actions that have been beneficial in the past, and after convergence, it selects the action with the maximum Q-value according to the principle of Q-learning. The definition of actions, states, and reward function in our approach depends on the task to be accomplished and the execution environment. An adaptive definition is provided in Section IV. The use of RL in this system is challenging because the final return concerns all agents, and each agent may question their share of the return. Additionally, since an agent cannot see their teammates' choices, they may be punished for a bad choice made by another agent. To address this issue, we consider all agents to be independent learners. The main challenge for these agents is coordination; how to ensure that all agents choose their individual actions consistently to achieve a Pareto-optimal joint action, where no other strategy benefits any of the agents. This is a complex problem resulting from the combined actions of several factors. Since agents must cooperate, they have no interest in threatening each other, but they must change their policy to improve their rewards and adapt to this change. We can model this setting as a repeated game where the same agents play the same game repeatedly. As the agents share the common goal of achieving the best end result, this game is cooperative and the reward is shared. A simple extension of centralized Q-Learning [56], case of a single learner, to stochastic games takes into account common actions in the calculation of Q-values. Thus, the update equation according to a centralized view of a system of n agents is:

$$Q(s, a_1, \dots, a_n) \leftarrow (1 - \alpha)Q(s, a_1, \dots, a_n) + \alpha^*[r + \gamma^* \max Q(s', a'_1, \dots, a'_1)] \quad (2)$$

Where s' is the new state, α is the learning rate and $\gamma \in [0, 1]$ is the discount factor [23].

In this model, the reinforcement perceived by an agent depends on the actions chosen by the group. Therefore, an agent does not know exactly its share in the total reward, or at least the influence of the received return (positive or negative) on its individual action. Even if an agent executes a good action, it could still be punished because of a bad choice made by the group. It is therefore preferable for an agent to give little importance to a punishment received after choosing an action that has satisfied it in the past. However, the agent must not be completely blind to sanctions, as this could result in a sub-optimal equilibrium or prevent coordination on an optimal joint action [29]. To solve this problem, we use the hysteretic Q-learning algorithm. This algorithm considers that an agent with an optimal individual action should not be punished because of a bad choice made by the group, but it must remain optimistic in order to reduce variations in the learned policy. The equation for updating the hysteretic Q-learning proposed by L. Matignon [29] for an agent i executing the action a_i from state s to transit to state s' is:

$$\delta \leftarrow r + \gamma \max_{a'} Q_i(s', a') - Q_i(s, a_i) \quad (3)$$

$$Q_i(s, a_i) \leftarrow \begin{cases} Q_i(s, a_i) + \alpha\delta & \text{if } \delta \geq 0 \\ Q_i(s, a_i) + \beta\delta & \text{otherwise} \end{cases} \quad (4)$$

Where α and β are two coefficients corresponding respectively to the increase or the decrease of the Q-value of a joint action. To have optimistic learners α must be greater than β . The main goals of using these two coefficients is to minimize the shadowed equilibria's effect and to manage stochasticity of the environment [29]. The Hysteretic Q-Learning algorithm is decentralized; each agent builds his own Q-table whose the size is independent on the number of agents and is linear according to his own actions. Algorithm 2 summarizes the hysteretic Q-learning.

Alg.2: Hysteretic Q-learning algorithm

Begin

Initialize arbitrarily $Q_i(s, a_i)$ for each (s, a_i) from $S \times A_i$

Initialize the initial state s

While s is not an absorbent state do

In the state s , choose the action a_i / phase of decision */*

Apply the action a_i and observe the new state s' and the return r

$q \leftarrow r + \gamma \max_{b \in A_i} Q_i(s', b)$

/ Hysteretic update */*

if $q \geq Q_i(s, a_i)$ then

$Q_i(s, a_i) \leftarrow (1 - \alpha)Q_i(s, a_i) + \alpha q$

else

$Q_i(s, a_i) \leftarrow (1 - \alpha)Q_i(s, a_i) + \beta q$

if $Q_i(s, \arg \max_{u \in A_i} \pi_i(s, u)) \neq$

$\max_{u \in A_i} Q_i(s, u)$ then

choose randomly $a_{max} \in \arg \max_{u \in A_i} Q_i(s, u)$

$\forall b \in A_i \pi_i(s, b) \leftarrow \begin{cases} 1 & \text{if } b = a_{max} \\ 0 & \text{otherwise} \end{cases}$

/ equilibria selection */*

$s \leftarrow s'$

End

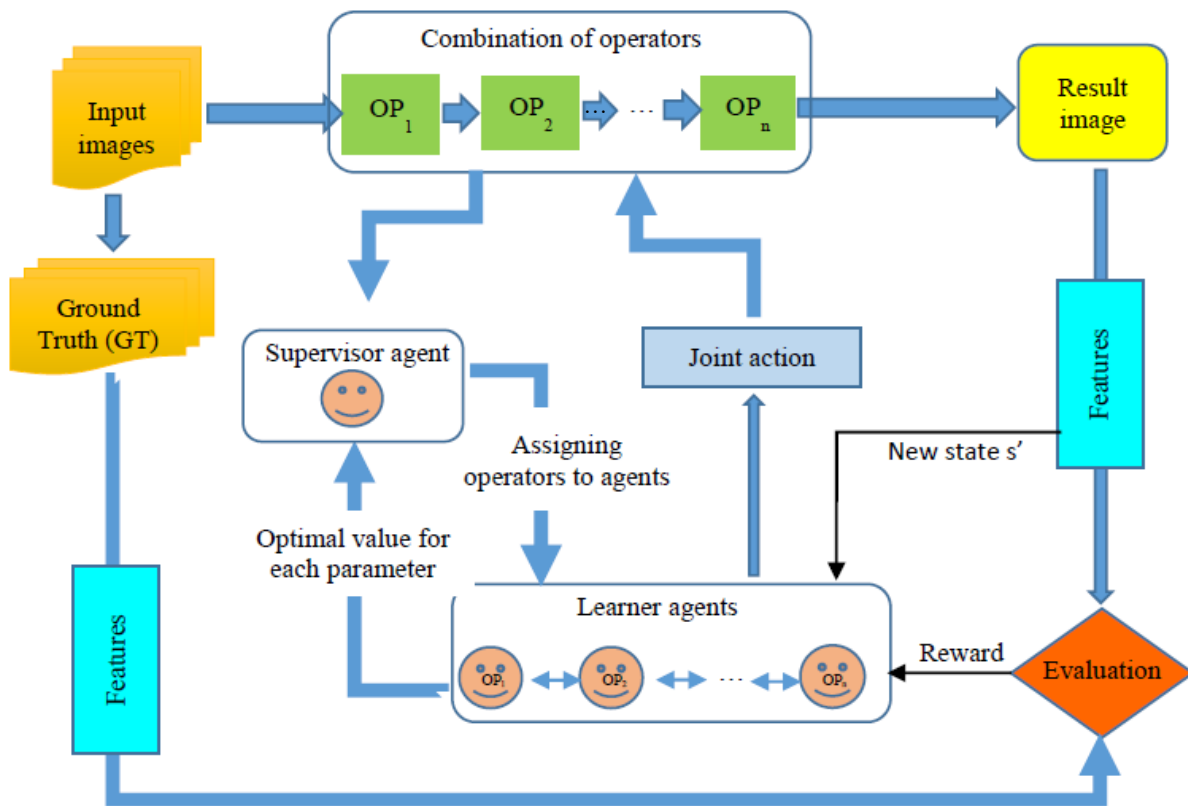


Fig. 3. Global schema of the proposed solution.

The proposed multi-agent system consists of two types of agents: learner agents and a supervisor agent. Each learner agent takes one operator and proceeds to estimate its parameters, with each agent representing one of the operators in the combination. The learner agents work together to learn the best values of the parameters for the entire combination of operators. These agents use a hysteretic Q-learning algorithm, which means that they give little importance to the penalties received but are not completely blind to them [37, 38].

The supervisor agent is responsible for distributing the operators to the learner agents and receiving the optimal value of each parameter.

V. EXPERIMENTS AND ANALYSIS

In this section, we evaluate the effectiveness of our approach by applying it to estimate parameters in an object recognition task. Object recognition is a vital area of computer vision with numerous applications such as object tracking, facial recognition, and autonomous driving. The main goal of object recognition is to detect objects in an image by locating, classifying, and framing them with rectangles. Object detection methods can be broadly categorized into two categories: the first approach divides the image into regions and classifies them into object categories, while the second approach treats object detection as a classification or regression problem, producing final results directly. Examples of methods in the first category include R-CNN [39], SPP-net [40], Fast R-CNN [41, 42], Faster R-CNN [43], R-FCN [44], FPN [45] and Mask R-CNN [46]. For the second category we find MultiBox [47], AttentionNet [48], G-CNN [49], YOLO

[50], SSD [51], YOLOv2 [52], DSSD [53] and DSOD [54]. While our work does not aim to propose a new object detection method, we demonstrate how our approach can provide valuable assistance to users seeking to determine the optimal values of each parameter in a combination of operators. We use object detection as a case study to demonstrate the effectiveness of our approach.

A. Experiments' Environment

We conducted an experiment to evaluate our approach using a dataset of 60 mixed textured images, where a disc (also textured) was inserted into 40 images while the remaining 20 did not contain it. The objective was to recognize and extract the disc from these images using a combination of two operators in two phases. In the first phase, a filter operator with two parameters needed to be estimated, and in the second phase, an operator with two parameters was applied to segment textures and classify them into clusters. For each parameter, a set of possible values was proposed, and a system of two agents was assigned to adjust the parameters for each operator. The results obtained using our proposed approach were compared with those found in [7], where the learning was centralized. To facilitate a comprehensive comparison, we implemented our approach using Matlab, which has a rich toolbox of image processing operators and allows for parallel programming using "Workers."

B. Parameter's Value Learning

The operators used in this experiment are "imfilter" and "GLCM_KMeansFct." The "imfilter" operator is an existing operator in the Matlab toolbox and is used for image filtering.

The "GLCM_KMeansFct" operator is a function that we implemented by combining two Matlab operators: "graycomatrix," which computes the gray-level co-occurrence matrix (GLCM) of an image, and "kmeans," which classifies the obtained textures into clusters.

The operator "imfilter" has two parameters:

- the type of filtering with a single value: {'unsharp'}
- alpha (smoothing coefficient) with 2 possible values: {0.2, 0.6}

The operator "GLCM_KMeansFct" has two parameters to adjust:

- the size of the sliding window with 7 possible values: {3, 5, 7, 9, 11, 13, 15}
- the number of clusters with 4 possible values: {2, 3, 4, 5}

The proposed approach in this paper consists of assigning one agent AG1 to "imfilter" and another AG2 to "GLCM_KMeansFct".

C. RL Configuration

The three principal components: actions, states and the return function must be defined adaptively to our approach.

1) *Actions*: An action of the proposed multi-agent system is a joining of individual actions of all the agents. For instance, an individual action for the agent AG1 is {unsharp, 0.2} or {unsharp, 0.6}, and for the agent AG2 is every combination between a size of the sliding window and a number of possible clusters. Thus, a joint action may be, for example, {unsharp, 0.2, 3, 2}.

2) *States*: A state is defined according to the features of the image obtained after the execution of an action. In this paper, we consider four characteristics to define a state.

$$S = [x_1, x_2, x_3, x_4] \quad (6)$$

x_1 : the number of objects in the result image.

x_2 : the ratio between the area of the result object and the area of the entire input image.

x_3 : the ratio between the area of the result object and the area of the reference object.

x_4 : the average of the values of the textural metrics: energy, correlation, entropy and contrast [55].

3) *Return function*: The return can be a punishment or a reward, depending on the quality criterion representing how well the object has been detected. A simple method is to use an objective assessment by comparing the obtained result with the ground truth. This comparison is made between features of the two images to generate a value determining a reward or a punishment. The value calculated from this comparison is a weighted sum of the difference between the features extracted from the two images. The weights reflect the importance of a feature in the final decision.

$$D = \sum_i w_i D_i \quad (7)$$

Where w_i are the weights assigned to each of the following differences:

D1: difference in number of objects;

D2: difference in size of objects;

D3: difference in area of objects;

D4: difference in values of entropy;

If D is greater than a given threshold, the return is a reward. Otherwise it is a punishment.

$$\text{if } (D \leq \varepsilon) r = +10; \\ \text{else } r = -10;$$

Where ε is the threshold. In this experiments, we fix it in 0.15.

D. Results

The experience is based on 40 textured images containing four different textures. In these images we inject an object of interest, which is a disk. Fig. 4 shows some examples of these images.

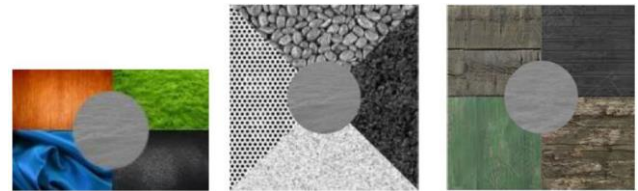


Fig. 4. Examples of used images.

The process of our approach needs a ground truth. In this experiment, the ground truth is the disc shown in Fig. 5.



Fig. 5. The ground truth in our experiment.

The multi-agent system we propose in this paper, uses some features to compare the obtained result with the ground truth. These features are the same for both images. The reference features are the following data:

NbrTargetArea = 5: represents the number of target zones

TotalSize = 20800: the total size of the image

AreaObjet = 2573: the area of the object of interest.

TextureMesure = 4.550863e+000: the entropy value (GLCM)

The desired state is [1.000 1.000 1.000 1.000] according to the equation (6).

The size and the area are measured in pixels.

For the process of exploration/exploitation, we use the hysteretic Q-learning with the values: Number of episodes and steps (iterations) are respectively 900 and 60, $\alpha = 0.5, \beta = 0.01, \gamma = 0.9, \varepsilon = 0.008$.

After running the proposed multi-agent system, we obtain the extracted object shown in Fig. 6.

The corresponding features of the result disc are:

NbrTargetArea = 6

TotalSize = 23810

AreaObjet = 2371

TextureMesure = 4.3448e+000



Fig. 6. Disc obtained by the proposed MAS.

To compare the two objects, the differences in equation (7) are calculated. Their obtained values are:

$D1 = 2.000000e-001$

$D2 = 1.447115e-001$

$D3 = 7.850758e-002$

$D4 = 2.06063e-001$

The difference between the result disc and the ground truth is given by:

$D = 0.2 * D1 + 0.2 * D2 + 0.3 * D3 + 0.3 * D4 = 9.8672e-002$

The optimal joint action (most rewarding) proposed by our MAS is then {'unsharp', 0.6, 5, 3} and its corresponding state is:

[1.1000 1.1000 0.9000 1.0000]

E. Discussion

The value of D is small; this means that the obtained disc is very closer to the reference. This result demonstrates that the proposed multi-agent system succeeds to extract the object of interest. The main contribution of this paper is not only to propose a method that finds the optimal parameters' values, that is already done in [7], but to propose an approach that outperforms the solution proposed in [7] in terms of speed and accuracy. Fig. 7 shows the curves of learning and the reward gain for the multi-agent system proposed in this paper.

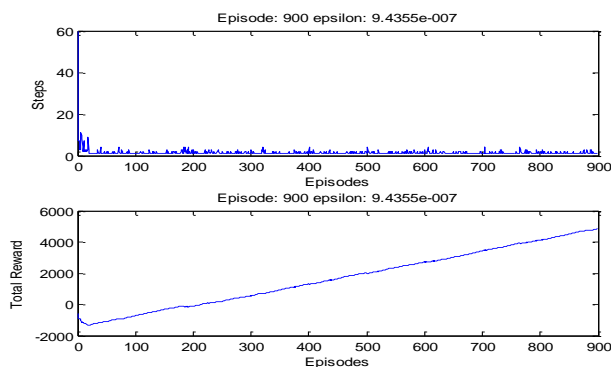


Fig. 7. Results of the proposed approach. Top: Learning curve (steps/episodes). Bottom: Corresponding cumulative reward.

The curve at the top shows the learning process. During the first 10 episodes, the multi-agent system executes several iterations to reach a convergence. A correspondence is clear in the curve below showing the cumulative returns. Indeed, during the first 10 episodes the multi-agent system receives only punishments, then the curve increases to show that the multi-agent system accumulates rewards. This is very logical with the principle of reinforcement learning, and in particular with hysteretic Q-learning.

To show the performance of the proposed multi-agent system as well as its main contribution, we run the one-agent approach [7] in the same environment with the same criteria. Fig. 8 shows the obtained curves.

In the one-agent approach, the top curve shows the execution of many iterations during the first 100 episodes with a high cumulative punishment as the bottom curve shows.

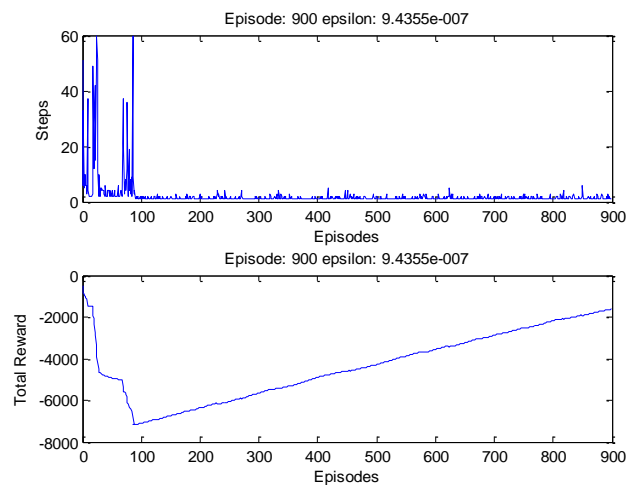


Fig. 8. Results of one-agent approach. Top: Learning curve (steps/episodes). Bottom: Corresponding cumulative reward.

Table I summarizes a comparison between the two approaches, by running them in the same environment and under the same conditions cited above.

TABLE I. COMPARISON BETWEEN MULTI-AGENT AND MONO-AGENT APPROACHES.

	Multi-agent & Hysteretic Q-Learning	One-agent & Q-learning
Quality Result (difference between the result and the ground truth)	10%	12%
Convergence (at which episode)	10	100
Reward (900 episodes)	4200	-
Punishment (900 episodes)	-1200	-7200

The results obtained using the multi-agent approach are significantly better than those obtained by the single-agent approach. The criteria used to evaluate the quality of the

extracted disc include the episode at which convergence starts, the values of punishment (which should be small), and the reward (which should be high). Based on these criteria, we can determine whether one approach is superior to another. In this paper, the multi-agent approach outperforms the single-agent approach. However, it is important to note that the difference between the extracted disc and the reference disc is still relatively high. This is mainly due to the choice of operators and not the proposed approach, which aims to adjust the parameters of the selected operators rather than propose a new method for object extraction.

VI. CONCLUSION

In this paper we have proposed an approach based on multi-agent system and reinforcement learning to automatize the process of parameter selection in image processing. In this work, adjusting parameters is seen as a decision process over time, where experiences gained from past decisions affect future decisions. The solution we have proposed, attributes one agent to an operator to adjust its parameters. We have used hysteretic Q-learning for multi-agent learning to find the best parameters for the given operators and test it to extract an object of interest. The complexity of the images, the speed of convergence and the quality of the results show the potential of the new approach and its adaptability. The results show that the proposed approach outperforms the use of one agent especially in terms of speed. Future works aim to include the operator selection also. In spite of using a predefined operator combination, we can suggest among several possible operators which are the best to use and furthermore what are their optimal parameters' values. Also, we think about using deep reinforcement learning instead of classical reinforcement learning.

REFERENCES

- [1] R. Deriche. Fast Algorithms for Low-Level Vision. *IEEE Transactions on Pattern Anal. and Machine Intell.*, vol. PAMI-12, no.1 (1990), 78-87.
- [2] E. Zemmour, P. Kurtser, and Y. Edan. Automatic parameter tuning for adaptive thresholding in fruit detection. *Sensors*, 19(9) (2019), 2130.
- [3] G. Rafael, M. Ricardo, A. Carlos, B. Péricles. Parameter tuning for document image binarization using a racing algorithm. *Expert Systems with Applications* 42 (2015), 2593–2603.
- [4] O. Miguel, P. Pablo, M. Otoniel, P. Manuel. Optimizing the image R/D coding performance by tuning quantization parameters. *Journal of Visual Communication and Image Representation* 49 (2017), 274–282.
- [5] P. Wang, F. Liu, C. Yang, X. Luo. Parameter estimation of image gamma transformation based on zero-value histogram bin locations. *Signal Processing: Image Communication*, (2018).
- [6] B. Diana, D. Michael, I. Dagmar. Global optimization using Gaussian processes to estimate biological parameters from image data. *Journal of Theoretical Biology* 481 (2019), 233–248.
- [7] I. Qaffou, M. Sadgal, A. Elfazziki. A New Automatic Method to Adjust Parameters for Object Recognition. *International Journal of Advanced Computer Science and Applications*, Vol. 3, No. 9 (2012), 213-217.
- [8] I. Qaffou, M. Sadgal, A. Elfazziki. Selecting Vision Operators and Fixing Their Optimal Parameters Values Using Reinforcement Learning. *Lecture Notes in Computer Science Volume 7340* (2012), 103-112. Springer-Verlag Berlin, Heidelberg ©2012.
- [9] I. Qaffou, M. Sadgal, A. Elfazziki. A Multi-Agents Architecture to Learn Vision Operators and their Parameters. *International Journal of Computer Science Issues*, Vol. 9, Issue 3, No 1 (2012), 140-149.
- [10] I. Qaffou, M. Sadgal, A. Elfazziki. Q-learning optimization in a multi-agents system for image segmentation. *International Journal of Advanced Studies in Computer Science and Engineering*, Volume 2, Theme based issue 3 (2013), 41-47.
- [11] F. Qingnan, C. Dongdong, Y. Lu, H. Gang, Y. Nenghai, C. Baoquan. Decouple Learning for Parameterized Image Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019.
- [12] J. Dong, L. Frosio, J. Kautz. Learning Adaptive Parameter Tuning for Image Processing, *Proc. EI, Image Processing: Algorithms and Systems XVI*, Burlingame (2018).
- [13] A.H. BOND and L.GASSER. Reading in distributed artificial intelligence. Morgan Kaufmann publishers, Inc, 1988.
- [14] J. FERBER. Les systèmes multi-agents: vers une intelligence collective, InterEdition. 1997.
- [15] S.J. Russell and P. Norvig. Artificial Intelligence: a Modern Approach. Prentice Hall, Englewood Cliffs, NJ, 2nd edition (2003).
- [16] P. Hofmann P. Lettmayer, T. Blaschke, M. Belgiu, S. Wegenkittl, R. Graf, T. Lampoltshammer, V. Andrejchenko. Towards a framework for agent-based image analysis of remote-sensing data. *International Journal of Image and Data Fusion*, 6:2 (2015), 115-137.
- [17] S.Y. Wai, C. WaiShiang, M. bin Khairuddin. Multi Agent Object Recognition: A Preliminary Study. *ICIIP* (2019), China.
- [18] T. Inguère, F. Carlier, V. Renault. Flexible image processing in Embedded Systems using Multi-agents Systems. *IFAC-PapersOnLine* Volume 49, Issue 25 (2016), 164-169.
- [19] A. Maudet, G. Touya, C. Duchêne, S. Picault. Patterns multi-niveaux pour les sma. *Journées Francophones sur les Systèmes Multi-Agents* (2015).
- [20] L. Males, D. Marcetic, S. Ribaric. A multi-agent dynamic system for robust multi-face tracking. *Expert Systems with Applications* 126 (2019), 246–264.
- [21] J. Paulin, A. Calinescu, M. Wooldridge. Agent-based modeling for complex financial systems. *IEEE Intelligent Systems*, 33 (2) (2018), 74–82.
- [22] R. S. Sutton and A. G. Barto. Reinforcement learning: an introduction. Adaptive computation and machine learning. MIT Press, Cambridge, Mass., 1998.
- [23] C. J. C. H. Watkins. Learning from Delayed Rewards. PhD thesis, Cambridge University, 1989.
- [24] S. Sehad. Contribution à l'étude et au développement de modèles connexionnistes à apprentissage par renforcement : application à l'acquisition de comportements adaptatifs. Thèse génie informatique et traitement du signal. Montpellier: Université de Montpellier II, 1996, 112 p.
- [25] E. Yang and D. Gu. Multiagent reinforcement learning for multi-robot systems: A survey. Tech. rep., Department of Computer Science, University of Essex, (2004).
- [26] M. Bowling and M. Veloso. An analysis of stochastic game theory for multiagent reinforcement learning. Tech. rep., Computer Science Department, Carnegie Mellon University, (2000).
- [27] B. Banerjee, S. Sen, J. Peng. On-policy concurrent reinforcement learning. *Journal of Experimental & Theoretical Artificial Intelligence*, 16(4) (2004), 245–260.
- [28] S. Abdallah and V. Lesser. A multiagent reinforcement learning algorithm with non-linear dynamics. *Journal of Artificial Intelligence Research*, 33 (2008), 521–549.
- [29] L. Matignon, J. G. Laurent, N. Le Fort-Piat. Independent reinforcement learners in cooperative Markov games: a survey regarding coordination problems. *Knowledge Engineering Review*, Cambridge University Press (CUP), 2012, 27 (1), pp.1-31.
- [30] L. Busoni, R. Babuska, B. De Schutter. A comprehensive survey of multiagent reinforcement learning. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on*, 38(2) (2008), 156–172.
- [31] C. Watkins and P. Dayan. Technical note: Q-learning. *Machine Learning*. 8 (1992), 279–292.
- [32] M. Lauer and M. Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *Proc. 17th International*

- Conf. on Machine Learning, pp. 535–542. Morgan Kaufmann, San Francisco, CA, (2000).
- [33] M. Bowling and M. Veloso. Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136 (2002), 215–250.
- [34] S. Kapetanakis and D. Kudenko. Reinforcement learning of coordination in heterogeneous cooperative multi-agent systems. In *AAMAS '04: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems*. Washington, DC, USA. IEEE Computer Society. (2004). pp. 1258–1259.
- [35] J.Hao, D. Huang, Y. Cai, H.-f. Leung. The dynamics of reinforcement social learning in networked cooperative multiagent systems, *Engineering Applications of Artificial Intelligence* 58 (2017), 111–122.
- [36] W. Zemzem and M. Tagina. Cooperative multi-agent reinforcement learning in a large stationary environment. *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, IEEE, 2017, pp. 365–371.
- [37] J. A. Swets. *Signal Detection Theory and Roc Analysis in Psychology and Diagnostics*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.
- [38] Chunyu, L., & Gang, L. Learning Multiple Instance Deep Representation for Objects Tracking. *Journal of Visual Communication and Image Representation* (2020), Volume 71, 102737.
- [39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR*, 2014.
- [40] K. He, X. Zhang, S. Ren, J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. PatternAnal. Mach. Intell.*, vol. 37, no. 9 (2015), pp. 1904–1916.
- [41] R. Girshick. Fast r-cnn. In *ICCV*, 2015.
- [42] Rossi L., Karimi A., Prati A. Self-Balanced R-CNN for instance segmentation. *Journal of Visual Communication and Image Representation* (2022), Volume 87, 103595.
- [43] S. Ren, K. He, R. Girshick, J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NIPS*, 2015, pp: 91–99.
- [44] Y. Li, K. He, J. Sun et al., “R-fcn: Object detection via region-based fully convolutional networks,” in *NIPS*, 2016, pp. 379–387.
- [45] T.-Y. Lin, P. Dollar, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie. Feature pyramid networks for object detection. In *CVPR*, 2017.
- [46] K. He, G. Gkioxari, P. Dollar, R. B. Girshick. Mask r-cnn. In *ICCV*, 2017.
- [47] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov. Scalable object detection using deep neural networks. In *CVPR*, 2014.
- [48] D. Yoo, S. Park, J.-Y. Lee, A. S. Paek, I. So Kweon. AttentionNet: Aggregating weak directions for accurate object detection. In *CVPR*, 2015.
- [49] M. Najibi, M. Rastegari, and L. S. Davis. G-cnn: an iterative grid based object detector. In *CVPR*, 2016.
- [50] J. Redmon, S. Divvala, R. Girshick, A. Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.
- [51] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A. C. Berg, Ssd: Single shot multibox detector,” in *ECCV*, 2016.
- [52] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger” arXiv: 1612.08242, 2016.
- [53] C. Y. Fu, W. Liu, A. Ranga, A. Tyagi, A. C. Berg. Dssd: Deconvolutional single shot detector. arXiv: 1701.06659, 2017.
- [54] Z. Shen, Z. Liu, J. Li, Y. G. Jiang, Y. Chen, X. Xue. Dsod: Learning deeply supervised object detectors from scratch. In *ICCV*, 2017.
- [55] R. M. Haralick, K. Shanmugan, I. Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 3, 1973, No 6, 610-621.
- [56] C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8: 279–292, 1992.

Design and Implementation of an IoT Control and Monitoring System for the Optimization of Shrimp Pools using LoRa Technology

José M. Pereira Pontón¹, Verónica Ojeda², Víctor Asanza³, Leandro L. Lorente-Leyva⁴, Diego H. Peluffo-Ordóñez⁵
Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador^{1,2}
SDAS Research Group, Ben Guerir, Morocco^{3,4,5}
Universidad UTE, Quito, Ecuador⁴
College of Computing, Mohammed VI Polytechnic University, Ben Guerir, Morocco⁵
Faculty of Engineering, Corporación Universitaria Autónoma de Nariño, Pasto, Colombia⁵

Abstract—The shrimp farming industry in Ecuador, renowned for its shrimp breeding and exportation, faces challenges due to diseases related to variations in abiotic factors during the maturation stage. This is partly attributed to the traditional methods employed in shrimp farms. Consequently, a prototype has been developed for monitoring and controlling abiotic factors using IoT technology. The proposed system consists of three nodes communicating through the LoRa interface. For control purposes, a fuzzy logic system has been implemented that evaluates temperature and dissolved oxygen abiotic factors to determine the state of the aerator, updating the information in the ThingSpeak application. A detailed analysis of equipment energy consumption and the maximum communication range for message transmission and reception was conducted. Subsequently, the monitoring and control system underwent comprehensive testing, including communication with the visualization platform. The results demonstrated significant improvements in system performance. By modifying parameters in the microcontroller, a 2.55-fold increase in battery durability was achieved. The implemented fuzzy logic system enabled effective on/off control of the aerators, showing a corrective trend in response to variations in the analyzed abiotic parameters. The robustness of the LoRa communication interface was evident in urban environments, achieving a distance of up to 1 km without line of sight.

Keywords—Control and monitoring system; shrimp pools; IoT architecture; LoRa technology; fuzzy logic control

I. INTRODUCTION

In recent years, there has been a remarkable growth in the aquaculture industry worldwide. One of the most prominent activities in commercial aquaculture is shrimp production. In this context, Ecuador has been a significant player due to its extensive coastline and longstanding tradition in shrimp farming. According to the 2020 Annual Report of the Instituto Nacional de Pesca (INP) of Ecuador [1], shrimp aquaculture is a strategic industry that contributes significantly to the country's economy.

Despite the favorable conditions for shrimp exports, not all shrimp farming companies have sufficient technology to meet the required care standards. This is because Ecuador faces limitations in technological development. The majority of

patents in the country are focused on preventing viral and bacterial diseases [2].

That is why the use of techniques such as Fuzzy Logic, which allows intelligent and adaptive control in complex systems, using linguistic rules to handle uncertainty and inherent imprecision in aquaculture, is of vital importance. According to [3], the use of Fuzzy Logic has been effective for control and intelligent management of water quality in aquaculture. In this work, they developed a simulation approach in MATLAB for a fuzzy logic-based control system for freshwater aquaculture. On the other hand, in [4], they present a review of works that employ fuzzy logic control in aquaculture systems, and in [5], supported by fuzzy logic and IoT, they develop an intelligent system for monitoring and early warning of water quality for aquaculture.

Additionally, the incorporation of LoRa as a long-range wireless communication technology provides the capacity to transmit data efficiently and reliably, even in remote areas. By combining these technologies with IoT, it is possible to create a real-time monitoring infrastructure and a centralized platform that allows shrimp farmers to supervise and control the shrimp harvest from any location.

Several works have been developed considering the aforementioned approach, such as the study conducted in [6], which highlights the importance of Long Range IoT technologies for remote monitoring and data transmission. On the other hand, in [7], they apply this technology to the monitoring of aquaculture information, and in [8], they explore the same area, developing a control and monitoring system based on IoT using LoRa. In the same vein, in [9], they design a system for monitoring water quality in aquaculture based on LoRa, obtaining encouraging results. On the other hand, the greatest emphasis that needs to be implemented in companies focuses on the shrimp breeding stage, as it is during this phase that the highest number of diseases and deaths in the harvest occur [10]. Shrimp diseases of infectious nature include viruses, bacteria, fungi, and parasites, where water quality directly influences the susceptibility of shrimp to different pathogens. For this reason, one of the most crucial considerations during shrimp breeding is the control of water quality [11].

TABLE I. IMPORTANCE OF ABIOTIC FACTORS

Parameter	Importance
Dissolved Oxygen (DO)	Dissolved oxygen produces crises of hypoxia or anoxia.
pH	When it is out of normal range it can cause stress.
Temperature	This parameter is related to DO in an inversely proportional way. It influences parameters such as solubility, chemical reactions, and toxicity.
Salinity	Affects the behavior of dissolved oxygen.
Turbidity	The cloudier the water, more light is blocked, affecting photosynthesis.
Ammonium	In shrimp farms it can be found as ionized (and non-ionized)

The factors affecting water quality are segmented into three categories: abiotic factors, biological factors, and environmental factors. Among these factors, we will emphasize the abiotic factors for the study, as they have a significant impact on production and harvest. In Table I, the importance of different abiotic factors in aquaculture is evident [12], [13], [14].

As can be seen in the previous table, the different abiotic factors have their influence on other abiotic factors, such as the behavior and growth of the shrimp. Therefore, it is necessary for the abiotic parameters to be within an ideal range.

The ideal range of abiotic parameters is determined based on different criteria presented by various authors. Table II presents the ideal ranges of abiotic parameters according to different authors.

TABLE II. ABIOTIC FACTORS' RANGES

Ref	Temp		Salinity		DO		pH		Nitrite		Ammonium	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
[14]	-	29	-	-	4.0	6.0	6.0	9.0	-	0.09	1.22	2
[15]	29	31	10	13	5.0	-	7.0	8.5	-	0.1	-	0.1
[16]	28	30	15	25	6.0	10.0	8.0	9.0	-	0.1	0.1	1.0
[17]	26	32	-	-	3.7	-	-	-	-	0.1	-	0.12
[18]	26	32	15	25	5.0	-	7.0	9.0	-	0.3	-	0.3
[19]	25	30	15	25	4.0	-	6.0	9.0	-	0.1	-	0.1

Having abiotic parameters within a defined ideal range is considered a controlled environment. There are different consequences when shrimp are not in a controlled environment, meaning when the abiotic parameters are either above or below the range.

To minimize the effect of abiotic factors' variation on aquatic life, monitoring and control of these factors must be carried out [20]. As the demand for shrimp continues to increase and there is a quest to enhance efficiency and sustainability in aquaculture practices, adopting advanced technologies for control and monitoring of shrimp pools becomes crucial. In response to this growing demand, the

present study develops a control and monitoring system based on Fuzzy Logic, LoRa (Long Range), and the Internet of Things (IoT) to address specific challenges faced by Ecuadorian shrimp farms. The approach combines these technologies to achieve intelligent and automated management of environmental conditions, maintaining an optimal environment for shrimp growth and development.

The rest of this manuscript is structured as follows: Section II presents a brief description of work related to the pond shrimp harvesting stage, abiotic factors, IoT architecture features, and the application of fuzzy logic. Section III describes the proposed methodology, the IoT structure, the devices to be used, the survey for the equipment selection and the control system. Section IV presents the experimental results obtained with the designed control and monitoring system. Finally, Section V provides the conclusions of this study and states the future work.

II. RELATED WORKS

The problem encountered during the shrimp pond harvesting stage is not unique to Ecuador; it is also present in various shrimp-producing countries. This section addresses how different authors have dealt with this challenge, and in Table III, a brief description of related works is provided.

TABLE III. RELATED WORKS

Reference	Description	Equipment
[12]	Proposal of a fuzzy model for analyzing the internal parameters of shrimp pools, describing the water status qualitatively.	-
[13]	Proposal for an analysis of water quality to control crises in aquatic systems.	-
[21]	Real-time modeling of a vehicle for shrimp pools, using fuzzy logic algorithms.	pH sensor. DO sensor. Temperature sensor. Turbidity sensor. Arduino Uno. Node MCU.
[22]	Evaluation of feeding strategies for shrimp based on fuzzy logic and mathematical functions.	Applied in a test laboratory. The simulation of the models was done using MATLAB.
[23]	Improving feeding conditions through the use of passive acoustics, computer vision, and telemetry.	Automatic feeder. Hydrophone. Controller. Wireless Communication.

In the work developed in [12], they established different categories for water quality based on abiotic parameters, which were obtained using fuzzy logic. From related literature, they determined that organisms are susceptible to diseases as a consequence of shrimp stress (variations in internal parameters of the pools). Fuzzy logic is considered an effective tool to assist shrimp farmers. In another research [13], it is mentioned that water quality is essential for proper shrimp growth, as they are susceptible to stress due to their ecosystem conditions. Other authors [14] indicate the importance of maintaining the ranges in which the abiotic parameters are found. These parameters are used to produce a quality index. On the other hand, in [21], they achieved a precision of 92% in the applied test to predict the water status. The control algorithm was based on fuzzy logic, and the authors anticipate that

implementing such elements in a shrimp pond will improve conditions. According to [23], having a better understanding of feeding techniques improves the conditions in which shrimp grow. Having a monitoring system in shrimp pools results in an increase in economic returns, as it allows better control of different parameters. The focus of that study was on feeding techniques, with a hydrophone being the selected equipment to analyze the ideal time for shrimp feeding.

A. IoT Architecture

An IoT architecture allows interconnection and communication between different devices by establishing a connection with the cloud. It is not necessary for the devices to be physically located in the same place; instead, the monitoring and visualization of various processes can be done from different platforms [24]. In a shrimp farm where access to technology is challenging, implementing IoT architecture concepts offers many advantages and benefits. Fig. 1 illustrates the characteristics that these services provide.

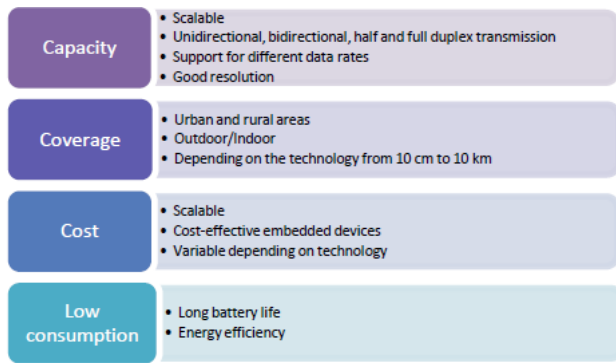


Fig. 1. Characteristics of IoT.

Wireless access networks within the IoT architecture vary and depend on the solution that will be provided to a specific problem, determining which one to use. Fig. 2 illustrates the coverage distance of different networks.

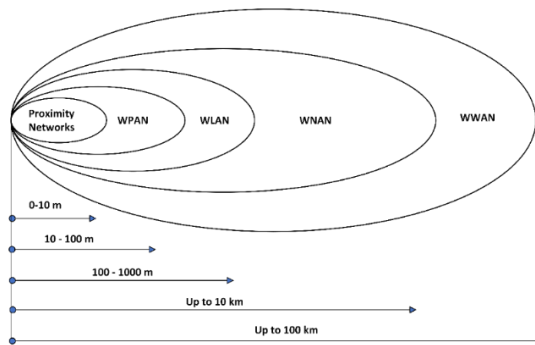


Fig. 2. Wireless access geographic coverage. Proximity networks, Wireless Personal Area Networks (WPAN), Wireless Local Area Networks (WLAN), Wireless Neighborhood Area Networks (WNAN) and Wireless Wide Area Networks (WWAN) [24].

The focus lies in the fact that different communication protocols have their limitations regarding their range. However, when a greater range is required, the use of Low-Power Wide Area Networks (LPWAN) within the can be employed. When considering the range that devices must have to communicate with each other, the modulation and

transmission rate must be carefully analyzed [25]. Table IV presents different communication protocols found in the literature, such as LoRa, Sigfox, and Zigbee [26].

TABLE IV. DIFFERENT COMMUNICATION PROTOCOLS

Characteristics	LoRa	Sigfox	Zigbee
Power	Low [27]	Low [28]	Low [29]
Transmission range	10 km	3 to 50 km	10 to 100 m
Data Rate	0.3 to 50 kbps	100 to 600 bps	20 to 250 kbps
Modulation	Spread spectrum modulation type based on FM pulses.	Ultra narrow band radio modulation	DSSS as a spreading technique
Modulation Technique	Chirp-spread spectrum	BPSK	BPSK
Topology	Star/mesh/point-to-point	Star	Star
Security	Resistance to electromagnetic interference. Robust to multi-path fading.	Low Frequency Accuracy constraint. High resistance to interferences.	Access control list. Frame Counters Encryption of over-the-air communications [30].
Battery	Long battery life	Long battery life	Long battery life
Bandwidth	900 MHz <500 kHz	200 kHz 100 Hz	2.4 GHz 915 MHz 868 MHz

Among the various advantages offered by different physical layer protocols, the LoRa modulation technique stands out for providing enhanced security. The Chirp Spread Spectrum, commonly employed by the military and in communication security applications, is utilized in LoRa [31].

III. MATERIALS AND METHODS

LoRa communications can be modified and depend on different parameters, which can be configured in the application. These parameters include spreading factor, coding rate, transmission power, chirp polarity, and synchronization word. Due to the versatility LoRa offers in device-to-device transmission, it has been selected as the communication method for the study. The LoRa protocol will be used for intercommunication between the different nodes in the pools. Meanwhile, the internet module will be used in the gateway module for communication with the ThingSpeak platform, enabling real-time visualization of the abiotic parameters of the different pools, as shown in Fig. 3.

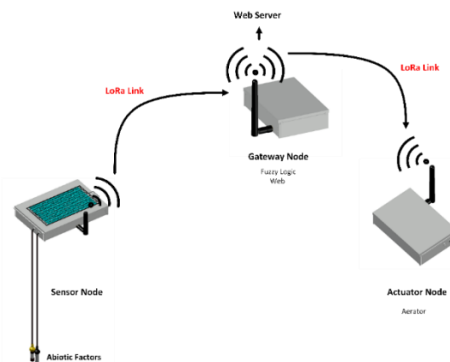


Fig. 3. Nodes.

In the gateway, the fuzzy logic function is embedded to calculate the next state of the aerator. The sensor nodes are enabled for data transmission, while the actuator nodes are enabled for data reception. The gateway node is configured to receive information from the sensor node and send information to the actuator node.

A. Control

In the process control field, there are different techniques for implementation, and one common type of controller is the PID controller, which provides a fast dynamic response. However, it requires control calibration for events with high precision, and one of its disadvantages is its high sensitivity to noise. An alternative control type is fuzzy logic, which consists of two stages: fuzzification and defuzzification (see Fig. 4).

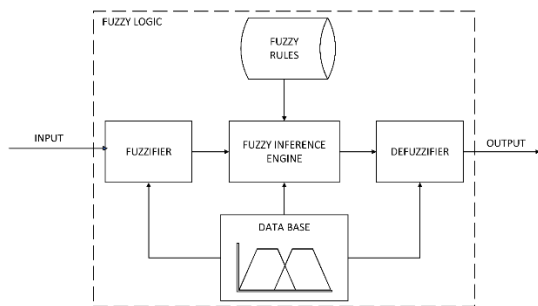


Fig. 4. Fuzzy logic.

The inputs to the controller first go through the fuzzification process to obtain a fuzzy value. This value then goes through an inference mechanism, which is complemented by fuzzy inference rules and fuzzy functions (database). Depending on the fuzzy input values, a fuzzy output value is obtained, which is then transformed into a real value through the defuzzification process.

B. Survey

For the selection of equipment to be implemented and sensors for the prototype, a survey was conducted with 35 shrimp farmers located in the province of El Oro, Ecuador. The most notable data from the survey were as follows:

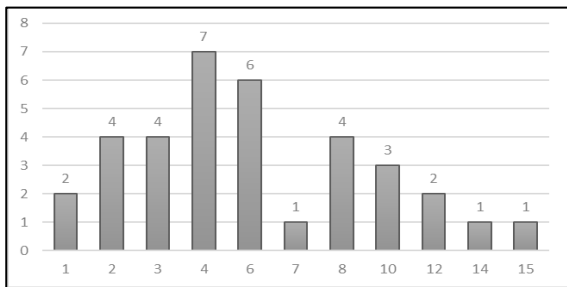


Fig. 5. Question 1: How many pools do you have?

The first finding pertains to the sizing of the pools and the scalability of the equipment. It is observed that the mode among shrimp farmers is to have four pools (Question 1). However, since there are shrimp farmers with 15 pools, the dimensioning of the devices to be connected in our monitoring network should be able to support and even have a larger capacity (see Fig. 5).

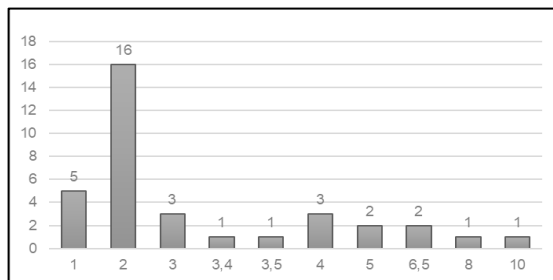


Fig. 6. Question 2: Size in hectares of your smallest pool

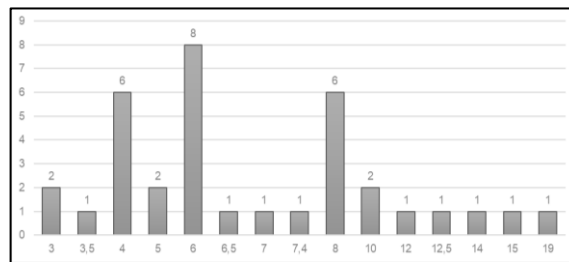


Fig. 7. Question 3: Size in hectares of your largest pool

Of the surveyed users, they were asked about the size of the smallest pools they have, and it was observed that they range from 1 to 10 hectares (Question 2) as shown in Fig. 6. Among these, 46% have a standard size of 2 hectares. When analyzing the case of the largest pools, they range from 3 to 19 hectares (Question 3) as depicted in Fig. 7.

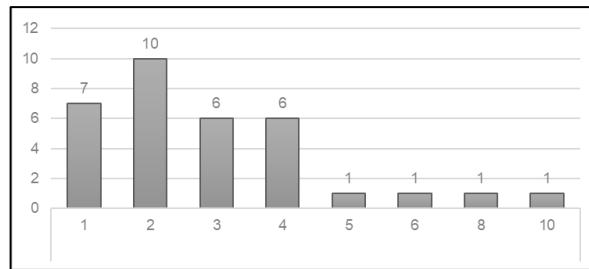


Fig. 8. Question 4: How many aerators do you have in your smaller pool?

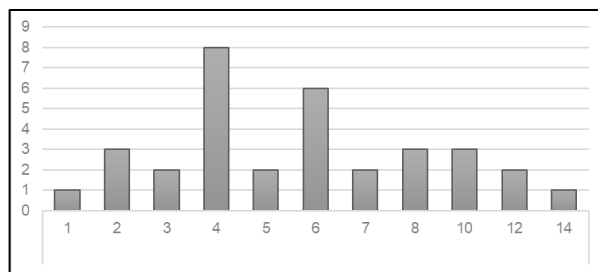


Fig. 9. Question 5: How many aerators do you have in your largest pool?

By knowing the size of the pools, we can identify how many devices or equipments are installed within each pool. It is observed that the range varies from 1 to 14 aerators (Questions 4 and 5 as shown in Fig. 8 and 9). Both points mentioned not only help determine how many pools are present but also give an idea of the number of devices that need to be installed and controlled in the pools.

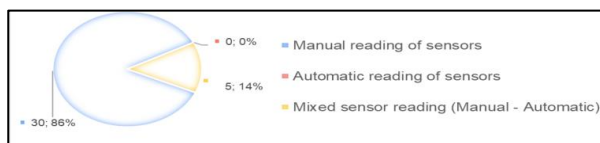


Fig. 10. Question 6: Indicate what kind of control mechanism you currently have implemented in your shrimp pool.

From the survey conducted with the shrimp farmers, it was determined that 30 of them perform the process of measuring parameters manually, while only five of them use both manual and automatic measurements (Question 6) as can be determined from Fig. 10.

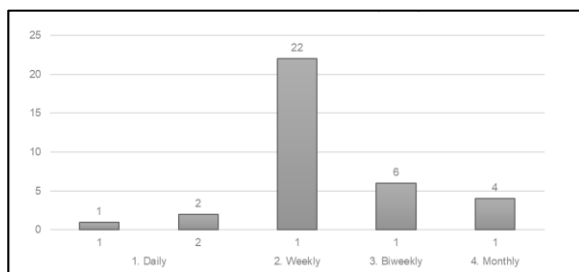


Fig. 11. Question 7: Measurement Frequency

A key question to determine the water quality among the surveyed shrimp farmers is that the measurement process is mostly done on a weekly basis, with only three shrimp farmers doing it daily. In the most severe cases, the analysis is done bi-weekly or monthly (Question 7) evident from Fig. 11. This information allows us to determine the focus of the final product to be developed.

C. Devices

For the selection of devices to be used, market devices were analyzed, and the summary is presented in Table V.

TABLE V. DEVICES

	Arduino Uno	Arduino Nano	TTGO Lora32 Oled V1	Heltec WiFi LoRa32
Chip	ATmega328P	ATmega328	ESP32	ESP32
Module	-	-	SX1276	SX1276
Processor	-	-	32-bit LX6	32-bit LX6
Transmission power	-	-	+20dB	+20dB
Transfer frequency	-	-	868-915 MHz	868-915 MHz
ROM	-	-	448 kB	448 kB
RAM	-	-	520 kB	520 kB
Flash Memory	16KB/32KB	32KB	4MB	8MB
Operating Voltage	5V	5V	2.7V-3.6V	3.3V-7V
Input Voltage	7V-12V	7V-12V	3.7V-4.2V	3.3V-7V
Input Voltage Limit	6-20V	-	-	-
GPIO	14	14	28	28
Analog Pins	6	8	-	-
PWM Pins	6	6	-	-
Clock Frequency	16 MHz	16 MHz	40MHz	40 MHz

In this way, by comparing the internal features of the devices, the TTGO LoRa32 Oled V1 model was selected, as it has a 32-bit processor and a 4MB flash memory. Although it has a smaller flash memory compared to the Heltec device, it still meets the needs for the sensor and actuator nodes. Having a low input voltage is favorable for achieving longer equipment autonomy.

D. Solution Description

For the monitoring and control of shrimp pools, the solution consists of a stage of measuring the abiotic parameters in the pond using a sensor node. Once the data is obtained, it is sent to a gateway node where it is processed using a fuzzy logic algorithm. This algorithm is designed to determine the next state of the pond's aerator (on or off). This value is then sent to an actuator node, which commands the activation of the aerator using a relay. A brief description of the process is shown in Fig. 12.

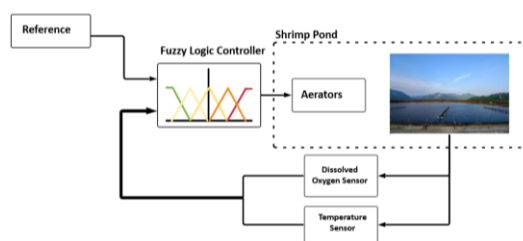


Fig. 12. Proposed control system.

The data obtained and used in this work are available at the following link: <https://github.com/josep5097/LoRa-Shrimp-Monitoring-Control>

As mentioned, the concept of the Internet of Things is used for the current model of the control and monitoring system. In Fig. 13, you can observe the structure of the proposed architecture.

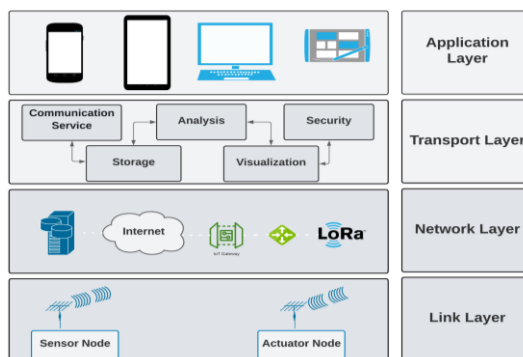


Fig. 13. IoT structure.

Within the link layer, there are sensor nodes and actuator nodes, and communication between the devices is through the LoRa protocol. When using the LoRa communication protocol, a topology must be chosen for the communication between devices. For the present system, a star network topology was selected over a mesh network. This decision was made because devices in a full duplex communication (mesh) require being powered on all the time, leading to a higher energy demand. An outline of the established architecture is shown in Fig. 14.

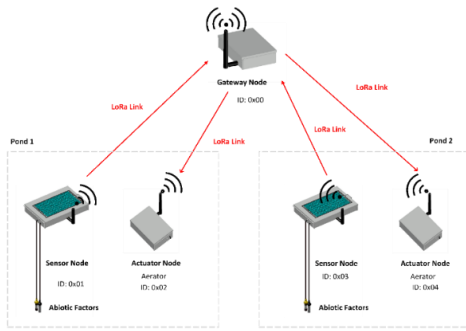


Fig. 14. Topology.

Within the network layer, there is a designated device to serve as our gateway to the Internet, in this case, the gateway node. It communicates with the different nodes in the pools using the LoRa protocol in its physical layer, and with the Internet using the HTTP protocol.

The devices have a synchronization word and ID, as shown in Fig. 13, which means that only devices with this synchronization word in their header can receive the message. Additionally, encryption was implemented to ensure that the messages transmitted by the LoRa devices are understood only by the intended receivers. To avoid signal interference, a gap between message transmissions must be established.

The captured data from the pools is visualized using the ThingSpeak platform, with the information being sent from the gateway device as described in Fig. 15.

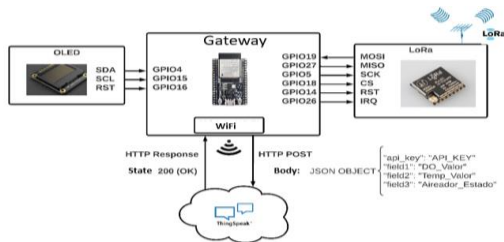


Fig. 15. Gateway connection.

In Fig. 15, besides the formation of the JSON object that is sent, you can see the pins used to connect both the OLED and the LoRa module. Similarly, the connection used with the sensor and actuator nodes is described. For the sensor node, GPIO pins are used for the dissolved oxygen and temperature sensors, as shown in Fig. 16.

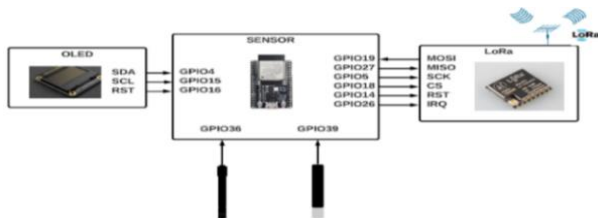


Fig. 16. Sensor node connection.

Meanwhile, for the actuator nodes, a GPIO pin is used to activate a relay for turning on or off an aerator (see Fig. 17).

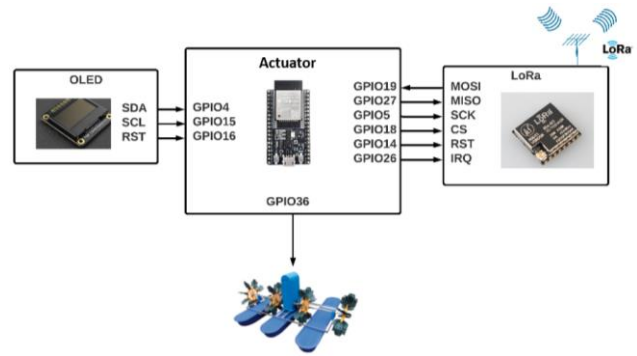


Fig. 17. Actuator connection.

E. Pseudocodes

To establish the described behavior of the devices in Fig. 15, 16, and 17, the processes are described using pseudocodes for each of the nodes.

Pseudocode 1: Nodo Gateway

```
# Init variables
DO [ ] = 0;
Temp [ ] = 0;
Aerators = 0;
# Init an array with all ID of nodes
ID_Sensors = {ID_S1, ID_S2};
ID_Actuators = {ID_A1, ID_A2};
# Init all variables in the fuzzy control
Fuzzy_Variables();
# Control Variables
num_Case = 1;
LoRa_Ok = false;

Void Setup {
    Fuzzy_Setup();
    LoRa.Setup();
    LoRa.begin();
    ThingSpeak.begin();
}

Void loop () {
    Switch (num_Case){
        case 1:
            LoRa_Read();
            If ( LoRa_Ok == True)
                num_Case ++;
            } else {
                num_Case = 1;
            }
            break;

        case 2:
            controlDifuso(Do, Temp,Aireadores);
            num_Case++;
            break;

        case 3:
            comunicacionLoRaActuador(destino, origen,
            Aireadores)
            num_Case++;
            break;

        case 4:
            procesoThingSpeak (Do, Temp, Aireadores,
            ID_Piscina);
            num_Case = 1;
            break;
    }
}
```

The monitored variables, such as DO and Temperature, are established along with the controlled variable, which is the Aerator. The Dissolved Oxygen and Temperature values will

be obtained from a sensor node, and the state of the Aerator represents the next state of the corresponding actuator node in the pool from where the sensor node's parameters originated. The value for the Aerator state will be determined using fuzzy logic. To establish a relationship between the sensor nodes and actuator nodes, unique IDs are assigned.

1) *Fuzzy logic configuration subprocess*: The fuzzy control process is described in Fig. 18, the diagram of relationships between entities, in which it can be observed how the variables interact among functions.

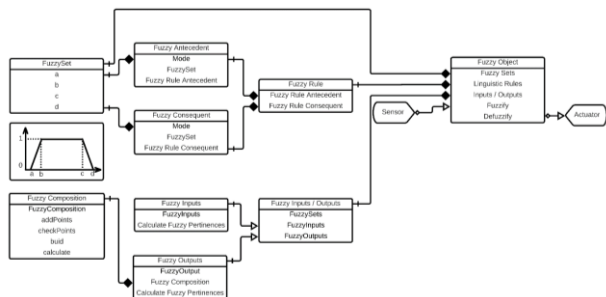


Fig. 18. Diagram of relationships between entities of fuzzy control.

In the fuzzy logic subprocess, the membership functions, as well as the fuzzy rules or linguistic rules, are established. Once these elements are set, the fuzzification process is applied to the inputs, which are values returned by the sensors, and the defuzzification process determines the value for the next state of the actuator.

2) *LoRa configuration subprocess*: The LoRa communication configuration consists of two stages:

- Internal parameter configuration (see Fig. 19).
- Message structure (see Fig. 20).

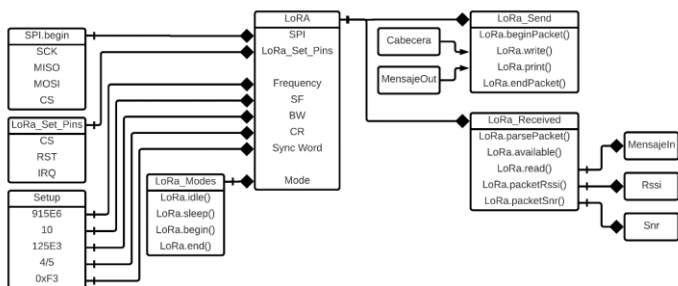


Fig. 19. Entity relationship of LoRa parameters.

For the configuration of internal parameters, the Entity Relationship of Fig. 19 is established, in which the same frequency, Spreading Factor (SF), Bandwidth (BW), Correction Rate (CR), and synchronization word (Init byte) are set for each device. The LoRa message sent has the same structure in different nodes, as shown in Fig. 20.



Fig. 20. LoRa message structure.

3) *Subprocess of thingspeak configuration*: Since the gateway node communicates with the cloud, it is necessary to establish the parameters for communication. This subprocess is described in the entity relationship of Fig. 21.

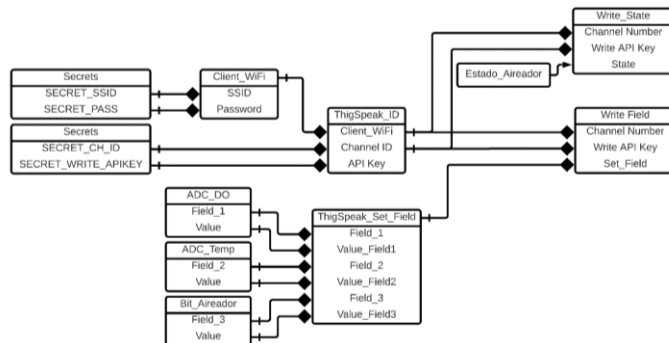


Fig. 21. Entity relationship of the thingspeak configuration subprocess.

The WiFi information, the channel to communicate with in ThingSpeak, and the API Key generated for communication are established. In this same scheme, it can be observed that when all the fields are ready and within the process flow, both a state is updated and fields are written to.

4) *Sensor node*: Since the sensor node is located inside the shrimp pond and operates autonomously, it requires better control over its operation time. Therefore, a reading interval of six hours is established for the process.

Pseudocode 2: Sensor node

```

Init
OD = 0;
Temp = 0;
ID_Coordinador = 0xID;
ID_Local = 0xID;
num_Case = 1;
Tiempo_A_Dormir = 21600; //21600 seconds = 6 hours
Void Setup {
Serial.begin(115200);
Display.begin();
LoRa.setup();
LoRa.begin();
}
Void loop {
Switch (num_Case){
case 1:
DO = lecturaAnalogicaPonderada(pinDO);
Temp = lecturaAnalogicaPonderada(pinTemp);
num_Case ++;
break;
case 2:
comunicacionLoRa(destino, origen, DO, Temp);
num_Case++;
break;
case 3:
num_Case=1;
esp_deep_sleep_start();
break;}}
    
```

5) *Actuator node*: Since the actuator node is directly powered by the control board, the equipment is always active. Its process involves reading the messages transmitted through a coordinator node, and depending on the case, it activates or deactivates the aerator.

Pseudocode 3: Actuator Node

```

Init
Aireador = 0;
ID_Coordinador = 0xID;
ID_Local = 0xID;
num_Case = 1;
Void Setup {
    Serial.begin(115200);
    Display.begin();
    LoRa.begin();
    pinMode(AireadorPin, OUTPUT);
}
Void loop {
    Switch (num_Case){
    case 1:
        LoRa_Read();
        If ( LoRa_Ok == True)
            num_Case ++;
        } else {
            num_Case = 1;
        }
        break;
    case 2:
        if (Aireador == 1){
            digitalWrite(aireadorPin, HIGH);
        } else {
            digitalWrite(aireadorPin, LOW);
        }
        num_Case = 1;
        LoRa_Ok = false;
        break;
    }
}
    
```

IV. RESULTS

In this section, the results obtained with the designed control and monitoring system are described. Fig. 22 illustrates the flow followed for the control of a shrimp pool.

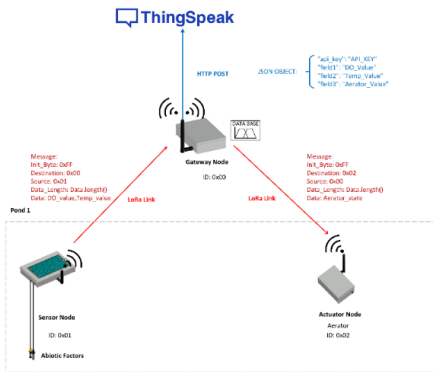


Fig. 22. Process of control in a shrimp pool.

In the embedded system of the coordinator node, linguistic terms, membership functions, antecedents, and consequents for fuzzy rules are established. In this study, three membership functions were applied, as shown in Fig. 23.

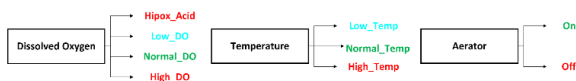


Fig. 23. Linguistic terms.

The embedded system uses fuzzy logic to represent the relevant linguistic terms. From these terms, three membership functions are generated: Dissolved Oxygen (see Fig. 24),

Temperature (see Fig. 25), and Aerator (see Fig. 26). These membership functions capture the characteristics and variability of each variable in the system.



Fig. 24. Dissolved oxygen membership function.

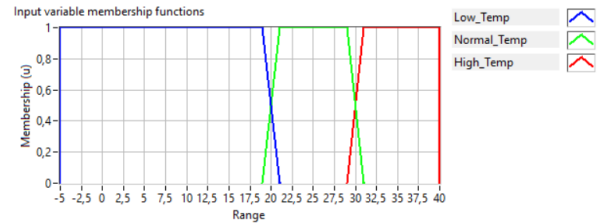


Fig. 25. Temperature membership function.

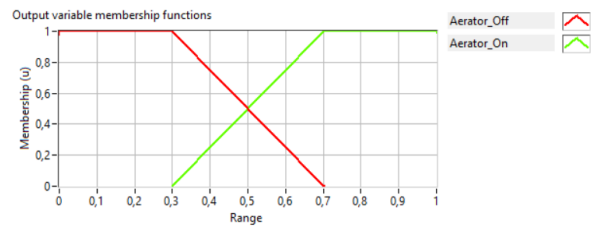


Fig. 26. Aerator membership function.

Since the fuzzy controller was implemented in an embedded system, the membership functions take a trapezoidal form, adapting to the limitations of the system. When establishing the membership functions, the values mentioned in Table II were taken into account. From these membership functions, the corresponding fuzzy rules were generated, which are shown in Fig. 27.



Fig. 27. Fuzzy Rules for the control system

The fuzzy rules used in this study were based on previous research [12], [13], [21]. From these established fuzzy rules, LabVIEW software was used to generate the fuzzy relation function, which reflects the expected behavior of the system. The fuzzy relation function was constructed based on the 12 defined fuzzy rules, and its response is shown in Fig. 28.

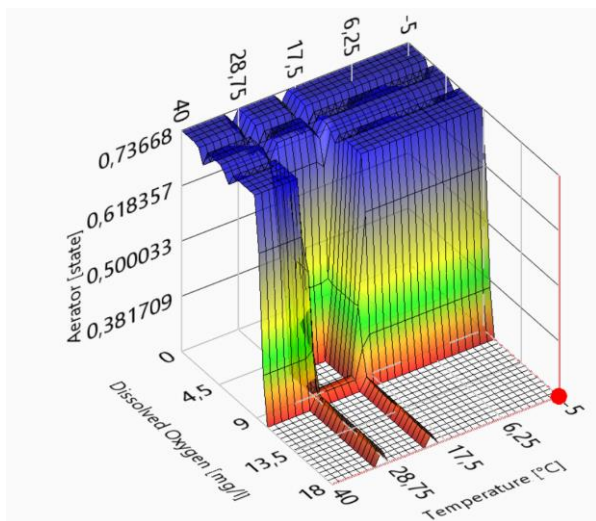


Fig. 28. Fuzzy rules.

In the coordinator node, a test was conducted with different values to verify the functionality, from which the results described in Table VI were obtained.

TABLE VI. FUZZY CONTROL RESPONSE

		Dissolved Oxygen (mg/L)												
		3.1	3.5	3.9	4.5	5	6	7	8	8.5	9.5	10	11	15
Temperature (°C)	-5	0.76	0.72	0.72	0.76	0.72	0.72	0.76	0.76	0.76	0.76	0	0	0
	10	0.76	0.72	0.72	0.76	0.72	0.72	0.76	0.76	0.75	0	0	0	0
	15	0.76	0.72	0.72	0.76	0.72	0.72	0.76	0.76	0.75	0	0	0	0
	20	0.76	0.72	0.72	0.76	0.72	0.72	0	0	0	0	0	0	0
	25	0.76	0.72	0.72	0.76	0.72	0.72	0	0	0	0	0	0	0
	30	0.76	0.72	0.72	0.76	0.72	0.72	0.65	0.65	0.65	0	0	0	0
35	0.76	0.72	0.72	0.76	0.72	0.72	0.76	0.76	0.75	0	0	0	0	

For the digital selection of the next state of the aerator, a threshold was established:

$$Aerator [state] > 0.5 \text{ then } Aerator \text{ ON}$$

Considering the threshold described previously, the table of the control response is modified (see Table VII).

The fuzzy system implemented in the embedded system achieves the performance demonstrated in Fig. 29. The interaction of the abiotic parameters with the mechanical movement of the aerator can be observed. The goal is that when the abiotic parameters are within an optimal range, the aerator switches to an off state.

TABLE VII. CONTROLLER RESPONSE - AERATOR STATUS

		Dissolved Oxygen (mg/L)												
		3.1	3.5	3.9	4.5	5	6	7	8	8.5	9.5	10	11	15
Temperature (°C)	-5	1	1	1	1	1	1	1	1	1	1	0	0	0
	10	1	1	1	1	1	1	1	1	1	1	0	0	0
	15	1	1	1	1	1	1	1	1	1	1	0	0	0
	20	1	1	1	1	1	1	0	0	0	0	0	0	0
	25	1	1	1	1	1	1	0	0	0	0	0	0	0
	30	1	1	1	1	1	1	0	0	0	0	0	0	0
35	1	1	1	1	1	1	0	0	0	0	0	0	0	

Fuzzy Logic System Response (ESP32 TTGO)

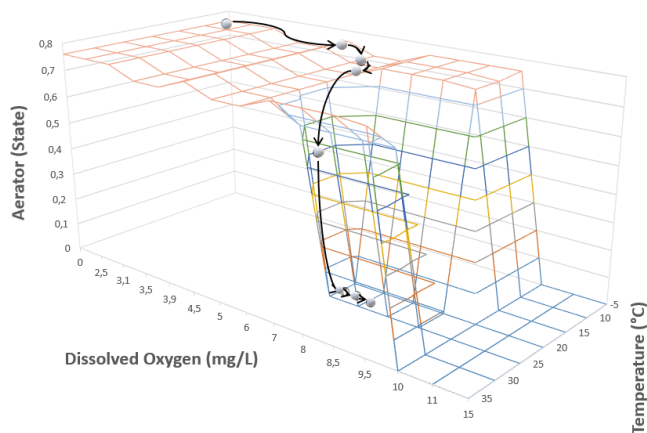


Fig. 29. Response of fuzzy logic according to abiotic factor.

V. CONCLUSIONS

The proposed control and monitoring system was established with the input of domain experts and reviewed research. However, it was observed that considering the needs of internal producers is of vital importance in order to encompass the required system activities and to determine the optimal solution based on available market equipment. It was determined that in order to operate within an IoT architecture, various situations and events must be taken into account. One of the key factors is the communication and security aspects that the equipment must possess, as well as the scalability they offer. Therefore, for the monitoring and control system established in this study, a star topology was considered for sensor and actuator nodes. This configuration allows for isolated points of connection, as establishing two-way communication requires higher processing and energy consumption.

Referring to the system's security levels, LoRa technology enabled the allocation of a specific channel, defining bytes for synchronization between devices. Additionally, fields were added to maintain control over transmissions and to determine the source and destination of messages. Identifying and discarding corrupted messages without processing them is crucial. To address this, a validation process was incorporated, which includes a field for assessing this condition.

The communication between the devices was successful in an urban environment of up to 1 km without packet loss, although it was necessary to adjust the LoRa configuration parameters to optimize performance. The use of fuzzy control based on abiotic factors allowed for effective control criteria in the shrimp pool, improving shrimp performance and growth through aerator control.

As future work, it is proposed to expand the number of abiotic factors without affecting performance and improve the range of the devices. An improvement option would be to migrate to a LoRaWAN network to take advantage of its additional benefits in terms of coverage and management capacity.

REFERENCES

- [1] Instituto Nacional de Pesca (INP), Estadísticas Acuícolas de Ecuador: Informe Anual 2020, Quito, Ecuador, 2021.
- [2] CEDIA, "Innovando el sector productivo," vol. 6, 2021.
- [3] S. D. Rana, and S. Rani, "Fuzzy logic based control system for fresh water aquaculture: A MATLAB based simulation approach," Serbian journal of electrical engineering, vol. 12, no 2, pp. 171-182, 2015.
- [4] K. Yue, and Y. Shen, "An overview of disruptive technologies for aquaculture," Aquaculture and Fisheries, vol. 7, no. 2, pp. 111-120, 2022.
- [5] H. C. Li, K. W. Yu, C. H. Lien, C. Lin, C. R. Yu, and S. Vaidyanathan, "Improving Aquaculture Water Quality Using Dual-Input Fuzzy Logic Control for Ammonia Nitrogen Management," Journal of Marine Science and Engineering, vol. 11, no. 6, 1109, 2023.
- [6] K. L. Tsai, L. W. Chen, L. J. Yang, H. J. Shiu, and H. W. Chen, "IoT based smart aquaculture system with automatic aerating and water quality monitoring," Journal of Internet Technology, vol. 23, no. 1, pp. 177-184, 2022.
- [7] M. Li, C. Lin, J. Ren, and F. Jiang, F, "A wireless ecological aquaculture water quality monitoring system based on LoRa technology," 2019 International Conference on Wireless Communication, Network and Multimedia Engineering (WCNME 2019), pp. 5-7, Atlantis Press, 2019.
- [8] H. Bates, M. Pierce, and A. Benter, "Real-time environmental monitoring for aquaculture using a LoRaWAN-based IoT sensor network," Sensors, vol. 21, no. 23, 7963, 2021.
- [9] S. -T. Chen, S. -S. Lin, C. -W. Lan and T. -I. Chou, "Design and development of a LoRa based Water Quality Monitoring System," 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien City, Taiwan, 2021, pp. 1-2.
- [10] J. Pereira, M. Mora, and W. Agila, "Qualitative Model to Maximize Shrimp Growth at Low Cost." 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM). IEEE, pp. 1-7, 2021.
- [11] N. Peña-Navarro, and A. Varela-Mejías, A, "Prevalencia de las principales enfermedades infecciosas en el camarón blanco Penaeus vannamei cultivado en el Golfo de Nicoya, Costa Rica," Revista de biología marina y oceanografía, vol. 51, no. 3, pp. 553-564, 2016.
- [12] J. J. Carbajal-Hernández, L. P. Sánchez-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Immediate water quality assessment in shrimp culture using fuzzy inference systems," Expert Syst. Appl., vol. 39, no. 12, pp. 10571–10582, 2012.
- [13] J. J. Carbajal-Hernández, L. P. Sánchez-Fernández, L. A. Villa-Vargas, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "Water quality assessment in shrimp culture using an analytical hierarchical process," Ecol. Indic., vol. 29, pp. 148–158, 2013.
- [14] N. C. Ferreira, C. Bonetti, and W. Q. Seiffert, "Hydrological and Water Quality Indices as management tools in marine shrimp culture," Aquaculture, vol. 318, no. 3–4, pp. 425–433, 2011.
- [15] M. Salah Uddin, M. Fatin Istiaq, M. Rasadin, and M. Ruhel Talukder, "Freshwater shrimp farm monitoring system for Bangladesh based on internet of things," Eng. Reports, vol. 2, no. 7, pp. 1–14, 2020.
- [16] C. E. Boyd, "Water quality in warmwater fish pond," Agric. Exp., p. 359, 1990.
- [17] B. H. Lam, H. X. Huynh, M. Traoré, P. Y. Lucas, and B. Pottier, "Monitoring environmental factors in mekong delta of vietnam using wireless sensor network approach," 8th Int. Conf. Simul. Model. Food Bio-Industry 2014, FOODSIM 2014, no. June, pp. 71–78, 2014.
- [18] N. C. Ferreira, "Aplicação de Índices de Qualidade de Água (IQA) como apoio à carcinicultura marinha," Ufsc, p. 62, 2009.
- [19] R. E. Yambay-Rueda and E. M. Alvarez-Alvarado, "Cultivo intensivo de camarón blanco litopenaeus vannamei en sistemas cerrados de recirculacion.," Universidad de Guayaquil, 2017.
- [20] M. G. Frías-Espericueta, M. Aguilar-Juárez, I. Osuna-López, S. Abad-Rosales, G. Izaguirre-Fierro, and D. Voltolina, "Los metales y la camaricultura en México," Hidrobiologica, vol. 21, no. 3, pp. 217–228, 2011.
- [21] K. Agustianto, T. Kustiari, P. Destarianto, and I. G. Wiryawan, "Development of realtime surface modeling vehicle for shrimp ponds (ReSMeV-SP)," IOP Conf. Ser. Earth Environ. Sci., vol. 672, no. 1, 2021.
- [22] R. A. Bórquez-Lopez, R. Casillas-Hernandez, J. A. Lopez-Elias, R. H. Barraza-Guardado, and L. R. Martinez-Cordova, "Improving feeding strategies for shrimp farming using fuzzy logic, based on water quality parameters," Aquac. Eng., vol. 81, no. May, pp. 38–45, 2018.
- [23] J. B. Darodes de Tailly, J. Keitel, M. A. G. Owen, J. M. Alcaraz-Calero, M. E. Alexander, and K. A. Sloman, "Monitoring methods of feeding behaviour to answer key questions in penaeid shrimp feeding," Rev. Aquac., vol. 13, no. 4, pp. 1828–1843, 2021.
- [24] C. Encinas, E. Ruiz, J. Cortez, and A. Espinoza, "Design and implementation of a distributed IoT system for the monitoring of water quality in aquaculture," In 2017 Wireless telecommunications symposium (WTS), pp. 1-7, IEEE, 2017.
- [25] B. S. Chaudhari, M. Zennaro, and S. Borkar, "LPWAN technologies: Emerging application characteristics, requirements, and design considerations," Futur. Internet, vol. 12, no. 3, 2020.
- [26] T. G. Durand, L. Visagie, and M. J. Booyesen, "Evaluation of next-generation low-power communication technology to replace GSM in IoT-applications," IET Commun., vol. 13, no. 16, pp. 2533–2540, 2019.
- [27] LoRa Alliance, "LoRa and LoRaWAN: Technical overview | DEVELOPER PORTAL." <https://lora-developers.semtech.com/documentation/tech-papers-and-guides/lora-and-lorawan/> (accessed Jun. 12, 2022).
- [28] Sigfox, "Technology | Sigfox." https://www.sigfox.com/en/what-sigfox/technology#id_security (accessed Feb. 12, 2022).
- [29] S. Farahani, "ZigBee/IEEE 802.15.4 Networking Examples," ZigBee Wirel. Networks Transceivers, pp. 25–32, 2008.
- [30] NXP, "Maximizing security in zigbee networks," 2017, [Online]. Available: <https://www.nxp.com/docs/en/supporting-information/MAXSECZBNETART.pdf>.
- [31] E. Aras, G. S. Ramachandran, P. Lawrence, and D. Hughes, "Exploring the security vulnerabilities of LoRa," 2017 3rd IEEE Int. Conf. Cybern. CYBCONF 2017 - Proc., no. June, 2017.

An Overview of Vision Transformers for Image Processing: A Survey

Ch.Sita Kameswari¹, Kavitha J², T. Srinivas Reddy³, Balaswamy Chinthaguntla⁴,
Senthil Kumar Jagatheesaperumal⁵, Silvia Gaftandzhieva⁶, Rositsa Doneva⁷

Department of Computer Science and Engineering (AI&ML), Keshav Memorial Institute of Technology, Hyderabad, India¹

Department of Information Technology, BVRIT HYDERABAD College of Engineering for Women, Hyderabad, India²

Department of Electronics and Communication Engineering, Malla Reddy Engineering College, Secunderabad, India³

Department of Electronics and Communication Engineering, Sheshadri Rao Gudlavalleru Engineering College,
Gudlavalleru, India⁴

Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi 626005, India⁵
University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria^{6,7}

Abstract—Using image processing technology has become increasingly essential in the education sector, with universities and educational institutions exploring innovative ways to enhance their teaching techniques and provide a better learning experience for their students. Vision transformer-based models have been highly successful in various domains of artificial intelligence, including natural language processing and computer vision, which have generated significant interest from academic and industrial researchers. These models have outperformed other networks like convolutional and recurrent networks in visual benchmarks, making them a promising candidate for image processing applications. This article presents a comprehensive survey of vision transformer models for image processing and computer vision, focusing on their potential applications for student verification in university systems. The models can analyze biometric data like student ID cards and facial recognition to ensure that students are accurately verified in real-time, becoming increasingly vital as online learning continues to gain traction. By accurately verifying the identity of students, universities and educational institutions can guarantee that students have access to relevant learning materials and resources necessary for their academic success.

Keywords—Vision transformers; image processing; natural language processing; image

I. INTRODUCTION

In recent years, deep neural networks such as convolutional neural networks (CNNs) [1], recurrent neural networks (RNNs) [2], graph neural networks (GNNs) [3], and attention neural networks [4] have been widely applied to a variety of artificial intelligence (AI) tasks. In contrast to previous non-neural models, which relied heavily on hand-crafted features and statistical methods, neural models can automatically learn low-dimensional continuous vectors as task-specific features from data, avoiding the need for complex feature engineering. Despite the popularity of deep neural networks, many studies have discovered that one of their fundamental limitations is their data-hungry nature. Due to many parameters in deep neural networks, they are prone to overfitting and have poor generalization capacity without appropriate training data [5].

CNNs are a fundamental component of modern computer vision systems. The advantage of CNNs was that they eliminated the need for manually constructed visual elements instead of learning to execute tasks “end to end” from data. The CNNs minimize manual feature extraction, and the CNN architecture is optimized for images and can be computationally expensive. Recent arguments have claimed that need goes beyond convolutions to represent long-range relationships. These initiatives aim to enhance convolutional models with content-based interactions, such as self-attention and non-local means, to improve performance in various vision tasks [6]. Transformers [7] are models that focus entirely on the self-attention process to establish global dependencies between input and output, and they have dominated natural language modelling in recent years [8-9]. Transformers and their variations have been thoroughly explored and used in natural language processing tasks such as machine translation [10], light-weight transformers [11], dynamic mask attention networks [12], language modelling [13], routing transformers [14], positional encoding schemes [15], and named entity identification [16-17]. The contrasts in size of visual elements and the high quality of pixels in images compared to words in text provide challenges in converting transformer from language to vision. The standard transformer is intended to process sequence data and is expected to receive a 1D series of token embedding. Many applications, including video understanding [18], image recognition [19], image super-resolution [20], object detection [21], segmentation [22], text-image synthesis [23] and visual question-answering [24], have been successfully implemented using transformer models and their variants in a variety of fields.

The survey in [25] explores recent advancements in visual transformers, an architecture originally designed for natural language processing but increasingly applied in computational visual media. The survey categorizes visual transformers based on task scenarios and analyzes their key ideas, with a particular focus on low-level vision and generation. The study reviews in detail backbone design approaches, offers quantitative comparisons, showcases image results, and includes information on computational costs and source code links to facilitate future development. In another recent survey by Jamil

et al. [26], the authors presented the first application of ViTs in computer vision, providing an overview of their usage and performance in various applications such as image classification, object detection, segmentation, compression, super-resolution, denoising, and anomaly detection, along with a comprehensive analysis of existing models, insights, and future research directions. Liu et al. [27] provided a comprehensive review of over one hundred visual transformers, attention-based encoder-decoder models inspired by the Transformer architecture in computer vision. It analyzes their effectiveness in fundamental tasks (classification, detection, segmentation) and different data stream types, presents a taxonomy to organize the methods, evaluates and compares them under various configurations, identifies unexploited aspects for further improvement, and suggests three promising research directions for future development. Subsequently, the survey [28] examines the advancements and trends in utilizing Transformers for video modeling, addressing their limitations with inductive biases and scalability. It analyzes how videos are handled at the input level, architectural modifications to enhance efficiency and capture temporal dynamics, various training regimes, self-supervised learning strategies, and provides a performance comparison against 3D ConvNets, demonstrating the superior performance of Video Transformers in action classification with reduced computational complexity.

Additional work in this approach may aid in a better understanding of Transformer models and detecting any erroneous behaviour or biases in the decision-making process. Since Transformer designs do not incorporate inductive biases (previous knowledge) to deal with visual input, transformers generally require a substantial quantity of training data in pre-training to determine the underlying modality-specific rules [29]. Several neural network architectures are known, including CNN, RNN, and transformer. CNNs were once the standard [30] in the Computer Vision domain, but transformers are gaining popularity [29]. While CNNs may capture inductive biases such as translation equivariance and localization, Vision Transformer overcomes inductive bias through large-scale training. According to the existing research [31], CNNs excel at small datasets, whereas transformers excel at massive datasets. The following fundamental issue is whether to employ in future CNN or a transformer.

Fig. 1 shows the number of publications on different image processing techniques using vision transformers [40].

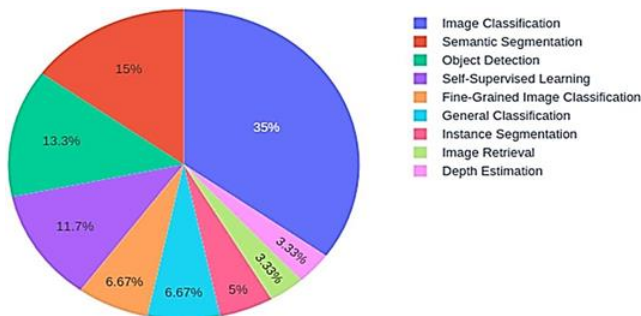


Fig. 1. The number of publications on different image processing techniques using vision transformers.

TABLE I. SUMMARY OF CONTRIBUTIONS FROM RECENT SURVEYS ON VISION TRANSFORMERS

Reference	Year	Scope	Contributions
Han et al. [32]	2022	General overview	Provides a comprehensive introduction to Vision Transformers
Chen et al. [33]	2021	Image classification	Focuses on the application of Vision Transformers for image classification tasks
Jamilet al. [26]	2023	General overview	Offers an in-depth analysis of Vision Transformers in various domains
Selvaet al. [28]	2023	Recent advancements	Highlights the latest research trends and advancements in Vision Transformers
Gehrig et al. [34]	2023	Object detection	Discusses the utilization of Vision Transformers for object detection tasks
Zhai et al. [35]	2022	NLP to computer vision transition	Explores the adaptation of Vision Transformers from natural language processing to computer vision
Yang et al. [36]	2022	Comprehensive review	Provides an extensive analysis of Vision Transformers and their applications -
Guo et al. [37]	2022	Comparison with CNNs	Compares the performance and characteristics of Vision Transformers with CNNs
He et al. [38]	2022	Medical image analysis	Examines the use of Vision Transformers in the field of medical image analysis
Aleissae et al. [39]	2023	Remote sensing applications	Surveys the application of Vision Transformers in remote sensing tasks

Table I summarizes significant contributions from the existing survey articles on vision transformers.

The contributions of this article are as follows:

- An overview of the background and preliminaries of vision transformers, widely used in natural language processing, and how they can be adapted for image processing.
- Discusses how vision transformers have been used for image classification and enhancement, which involves improving the quality of images by removing noise, enhancing contrast, and increasing resolution.
- Explores how vision transformers can be used for object detection, which is the process of identifying and locating objects within an image, and how they can achieve state-of-the-art performance on this task.
- Highlights the role of vision transformers in education and university systems, specifically in student verification, where they can automate the process of verifying student identities, making it faster and more accurate.
- Discusses how vision transformers can deal with multimodal tasks, where they can process and fuse information from multiple modalities, such as text, image, and audio.

The rest of this article is organized as follows. The second part introduces the background details of the transformer, and the third part explores the usage of the visual transformer variants. The fourth part throws light on multimodal variants of

using vision transformers. It also discusses the future research directions of visual transformers and the fifth part concludes the paper.

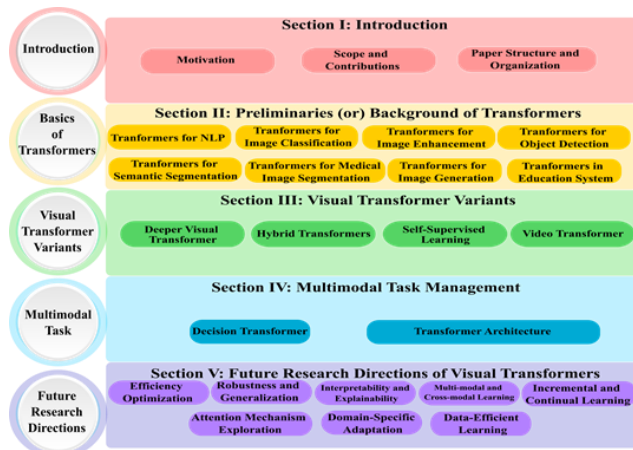


Fig. 2. Overall organization of the article.

Fig. 2 shows the overall organization of the sections presented in this article.

II. PRELIMINARIES (OR) BACKGROUND OF TRANSFORMERS

The transformer is made of L layers, each of which has two major blocks: a Multi-Headed Self Attention (MSA) layer that performs a self-attention operation on various projections of the input tokens and a Multi-Layer Perceptron (MLP). Both the MSA and MLP layers are preceded by layer normalization and followed by a skip connection [41]. The attention mechanism was initially applied for 1-D data processing in natural language processing [42-43]. It has recently expanded to handle two-dimensional images and three-dimensional video data [44].

The fundamental components of a transformer include Multihead Self-Attention (MSA), Multi-Layer Perceptron (MLP), and Layer Normalization (LN) [7]. The authors [45] proposed the Gaussian Error Linear Unit (GELU) used to a great extent as one of the high-performing activation functions for neural networks. The work in [46] discusses the challenges of applying batch normalization to RNNs and introduces a new technique (called layer normalization) which addresses these challenges. Layer normalization computes mean and variance for normalization from all summed inputs to neurons in a layer on a single training case. It is effective in stabilizing hidden state dynamics in recurrent networks. It significantly reduces training time compared to previous techniques.

A. Transformers for NLP

In recent years, the transformer has evolved into a fundamental component of numerous cutting-edge natural language processing (NLP) models. Like RNN, the transformer is a robust performance model helpful for standard NLP applications such as intent identification in a search engine, text creation in a chatbot engine, and classification. The authors proposed a feed-forward network design that relies entirely on attention processes and avoids convolutions and recurrence. It achieved state-of-the-art performance on several tasks significantly and generalized exceptionally well to other

NLP tasks, even with limited data. This design served as the foundation for numerous NLP models. GPT [47-49] and BERT [8] are two pioneering Transformer-based pre-trained models (PTMs) that employ autoregressive and autoencoding language modeling as pre-training objectives, respectively. Different Pre-trained models XLNet [50], RoBERTa [51], ALBERT [52], and T-NLG [53] are used in NLP tasks. Fig. 3 shows the structural difference between Transformer, GPT, and BERT [54].

Devlin et al. utilized the Transformer encoder (and only the encoder) to pre-train deep bidirectional representations from the unlabeled text. This pre-trained BERT model is fine-tuned with just one extra output layer to reach state-of-the-art performance for various NLP tasks without significant task-specific architectural changes. GPT [47] is a framework and training technique for natural language processing problems based on the Transformer architecture. The process for training is twofold. First, unlabeled data was used to learn the initial parameters of a neural network model using a language modeling aim. Then, using the associated supervised goal, these parameters are modified to a target task.

B. Vision Transformers for Image Classification

There have been many efforts to apply Transformers to vision tasks. These works are divided into two categories. The first category comprises models of pure attention. These models frequently use self-attention and strive to create convolution-free vision models. The second category encompasses networks developed using self-attention and convolutions [55]. Self-attention networks have revolutionized NLP and rapidly advanced image analysis tasks such as image classification and object recognition [56-57].

In computer vision, attention is employed in conjunction with or instead of CNN. This reliance on CNN is not required, as a pure transformer applied straight to sequences of image patches can do quite well on image classification tasks. The original text Transformer accepts a series of words as input and then uses them for classification, translation, or other natural language processing tasks. Dosovitskiy et al. [58] made the fewest feasible changes to the Transformer architecture to work directly on images rather than words for the vision transformer. Fig. 4 shows the architecture of the vision transformer.

Vision transformer generates a grid of square patches from an image. Each patch is converted to a single vector by concatenating the channels of all its pixels and then linearly projecting it to the chosen input dimension. Because transformers are structure-independent, they can add learnable position embedding to each patch, allowing the model to learn about the structure of the images. Vision transformer does not know the relative location of patches in the image or even if the image has a two-dimensional structure a priori. It must learn this information from training data and encode it in the position embeddings. Feed the sequence as an input to a state-of-the-art transformer encoder. Pre-train the vision transformer model with image labels, fully supervised then on an extensive dataset. Fine-tune the downstream dataset for image classification.

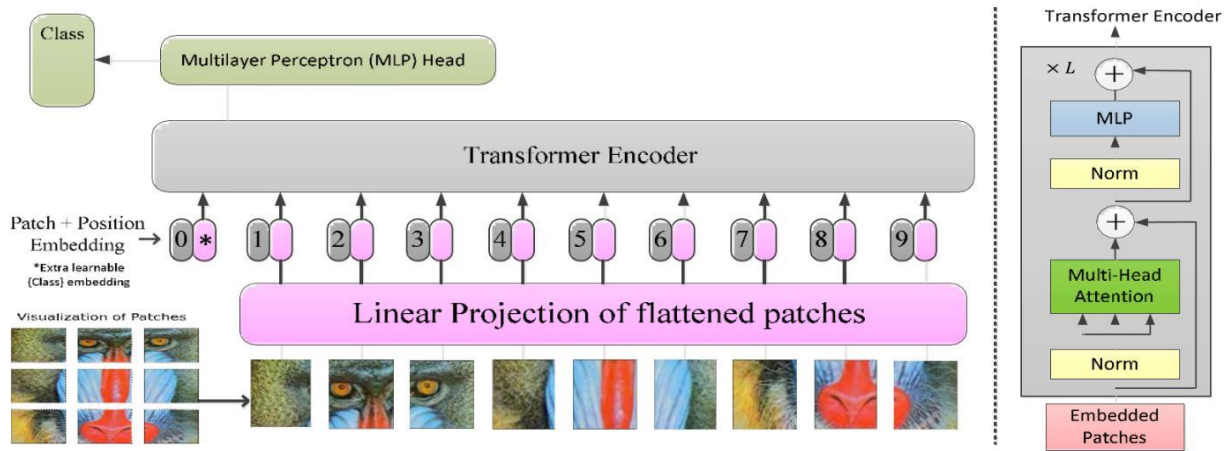


Fig. 3. Structure of transformer, GPT, and BERT.

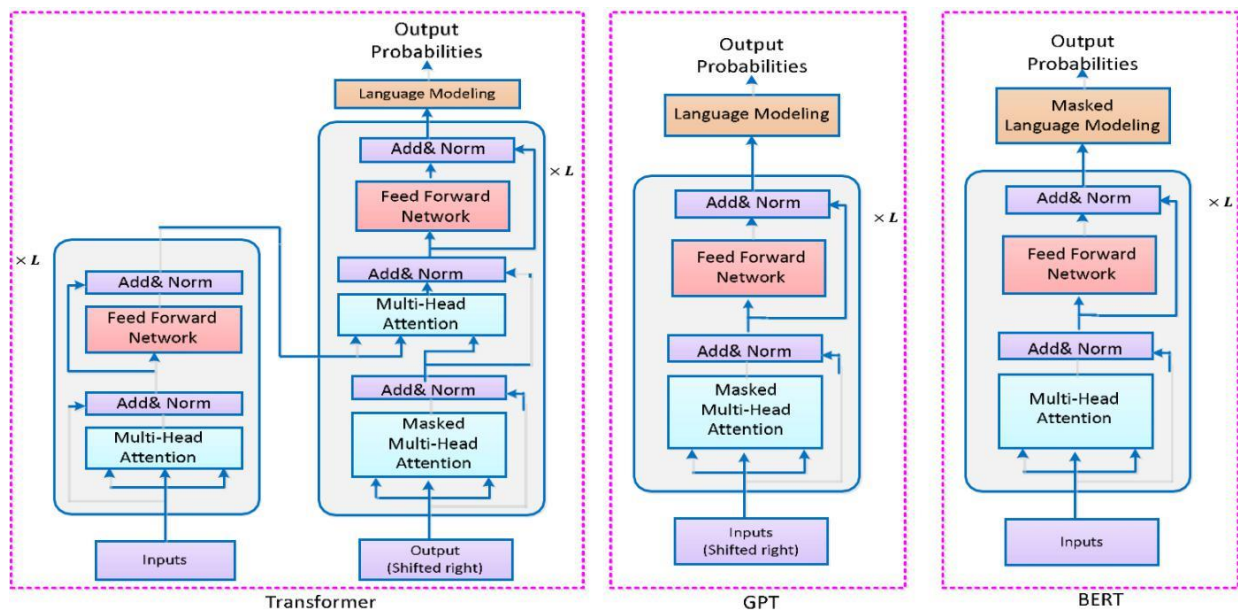


Fig. 4. Vision transformer architecture.

The Image GPT (iGPT) method [59] is an unsupervised generative pre-training technique for developing robust visual representations. By directly applying a GPT-2 [48] model to the image pixels, compelling image completions and samples were obtained, demonstrating that a completely Transformer-based architecture is viable for some visual tasks, regardless of the input image quality.

It does not need to prepare a large dataset to properly train the network in Data-efficient Image Transformers (DeiT) [6]. Instead, student-teacher setup and more intensive data augmentation and regularization are employed, such as stochastic depth [60] or repeated augmentation [61]. The teacher is a neural network designed to guide its student induction bias for convolutions [19].

LeViT is based on the architecture of the vision transformer [58] and the training technique of DeiT [19]. Regarding the speed/accuracy trade-off, LeViT considerably outperforms previous convnets and vision transformers [62]. LeViT is five times faster than EfficientNet on the CPU at 80.

To improve image classification accuracy, Chen et al. [51] describe Cross Vision Transformer (CrossViT) [63], a dual-branch vision transformer learning multi-scale features. The proposed technique analyses separately small-patch and large-patch tokens using two distinct branches with varying computational costs. These tokens are subsequently merged numerous times repeatedly by attention to complement one another. It also created an efficient token fusion module based on cross-attention using a single token for each branch as a query to exchange information with other branches. Cross-attention needs linear time for computational and memory complexity when it usually requires quadratic time.

Transformer-iN-Transformer (TNT) [64] combines both patch-level and pixel-level representation by utilizing an outer Transformer block that processes patch embedding and an inner Transformer block that models the relation between pixel embedding.

C. Transformer for Image Enhancement

Chen et al. developed a pre-trained image processing model based on the transformer design, namely *Image Processing Transformer (IPT)* [65]. The IPT model features multiple heads, multiple tails, and a standard transformer body for performing various image processing tasks such as super-resolution and denoising. The IPT model was trained using supervised and unsupervised methods, demonstrating a significant capacity to capture intrinsic characteristics for low-level image processing. Experiments indicate that IPT can outperform state-of-the-art techniques using a single pre-trained model following a brief fine-tuning phase.

Yang et al. [20] proposed a novel *Texture Transformer Network for Image Super-Resolution (TTSR)* in which the low-resolution (LR) and high-resolution (Ref) images are expressed as queries and keys, respectively, in a transformer. TTSR is a collection of closely linked modules designed for image generation tasks, comprising a learnable texture extractor based on deep neural networks, a relevance embedding module, a hard-attention module for texture transfer, and a soft-attention module texture synthesis.

D. Transformer for Object Detection

Carion et al. [21] introduced *DEtection TRansformer (DETR)* to eliminate the requirement for such hand-crafted components and developed the first fully end-to-end object detector with highly competitive performance. DETR is a basic architecture shown in Fig. 5 that combines CNNs with Transformer encoder-decoders [66]. They use Transformer's versatile and robust relation modelling capabilities to substitute hand-crafted rules when appropriately prepared training signals are used. DETR is a novel approach to object recognition based on transformers and bipartite matching loss for direct set prediction. Applied to the problematic COCO dataset, the method obtains results equivalent to an improved Faster R-CNN baseline. DETR is simple to construct and offers a modular design easily extendable to panoptic segmentation, resulting in competitive performance. Additionally, it outperforms Faster R-CNN on big objects, most likely because of the global information processing produced by self-attention.

However, it has its own range of difficulties. These difficulties are primarily due to the Transformer's attention deficiencies in handling image feature maps as essential elements: (1) DETR's ability to identify small objects is relatively poor. Modern object detectors use high-resolution feature maps to identify small objects more accurately. However, high-resolution feature maps would impose an excessive complexity level on the self-attention module of DETR's Transformer encoder, which scales quadratically with the spatial dimension of the input feature maps. (2) Compared to current object detectors, DETR takes more training epochs to converge. DETR is primarily due to the difficulty of training the attention modules that analyze visual characteristics.

Deformable (DETR) [21] remove the requirement for several handmade components in object detection while exhibiting acceptable performance. However, because of the limitations of Transformer attention modules in processing visual feature maps, it has a sluggish convergence rate and a restricted feature spatial resolution. To address these concerns, authors [67] suggested Deformable DETR, in which the attention modules focus exclusively on a limited number of critical sampling points surrounding a reference. Deformable DETR is a technique for object detection that seeks to address DETR's delayed convergence and high complexity problems. It combines the advantages of deformable convolution sparse spatial sampling with the relation modelling capability of transformers. Deformable DETR introduced a deformable attention module that uses a few sample sites as a pre-filter for conspicuous key components among all feature map pixels. Without relying on FPN, the module may be organically expanded to aggregate multi-scale characteristics. With ten fewer training epochs, deformable DETR can outperform DETR (particularly on tiny objects). Extensive trials on the COCO [68] benchmark validate this method.

Zheng et al. propose a novel transformer variation called the Adaptive Clustering Transformer (ACT) to reduce the computation cost associated with high-resolution input [69]. ACT uses Locality Sensitive Hashing (LSH) to cluster query characteristics adaptively and approximates the query-key interaction using the prototype-key interaction. ACT is capable of reducing the quadratic complexity inherent in self-attention.

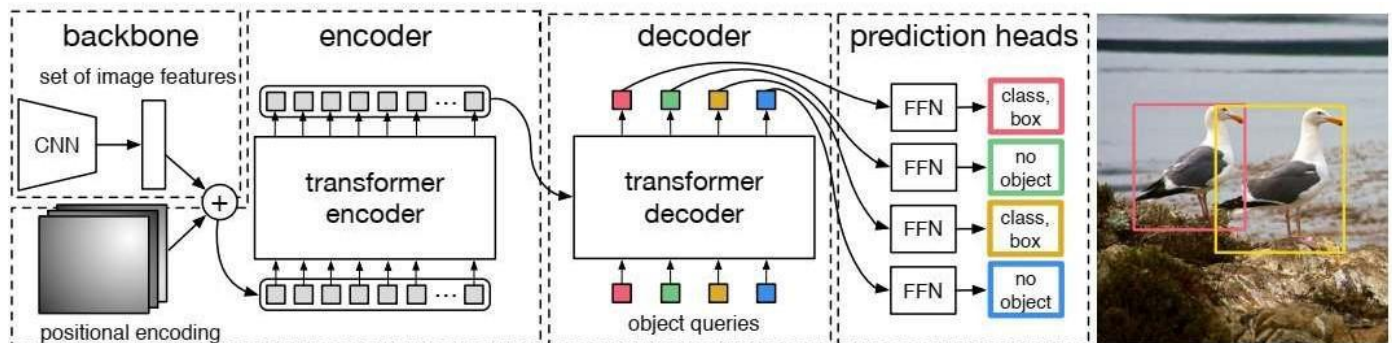


Fig. 5. DETR's general architecture. The image is from [21].

Inspired by the tremendous success of pre-training transformers in natural language processing, Dai et al. [70] proposed Unsupervised Pre-train DETR (UP-DETR) for object detection. The proposed UP-DETR model includes pre-training and fine-tuning procedures: (a) the transformers are unsupervised trained on a large-scale dataset without human annotations, and (b) the complete model is fine-tuned using labelled data, similar to the original DETR. Precisely clip random regions from the provided image and send them to the decoder as queries. Pre-trained on these query patches from the original image, the model detects them. The authors solve two critical difficulties during pre-training: multitask learning and multi-query localization. To balance classification and localization preferences in the pretext task, freeze the CNN backbone and propose a patch feature reconstruction branch optimized for conjunction with patch detection. (2) To accomplish multi-query localization, expand UP-DETR from single-query patches to multi-query patches by including object query shuffling and an attention mask. To expedite DETR's training convergence and prediction capability in object detection, Sun et al. [71] conduct extensive experiments and suggest two innovative methods, namely TSP-FCOS (transformer-based Set Prediction with FCOS) and TSP-RCNN (transformer-based Set Prediction with RCNN). These techniques converge considerably quicker than the original DETR and significantly outperform DETR and other baselines regarding detection accuracy

Beal et al. [72] developed ViT-FRCNN, a competitive object detection solution that uses a transformer backbone, implying that sufficiently distinct architectures from the well-studied CNN backbone are viable for advancement on complex vision problems. Transformer-based models have proven capacity to pre-train with large datasets without reaching saturation and fast fine-tune to different tasks, both observed with ViT-FRCNN.

The authors [73] suggested a novel variation of Vision Transformer models based on focal attention, called Focal Transformer that outperforms state-of-the-art (SoTA) vision Transformers on various publicly available image classification and object detection benchmarks.

Extracting strong feature representations is a significant issue in the re-identification of objects (ReID). Although techniques based on CNNs have gained considerable success, they analyze just one local area at a time and suffer from information loss due to convolution and down-sampling operators. To address these constraints, He et al. [74] introduced Transformer for Object re-identification (TransReID), a pure transformer-based object Reid framework. First, encode each image as a sequence of patches and then construct a transformer-based strong baseline with a few essential enhancements that obtain competitive performance on many Reid benchmarks using CNN-based techniques.

E. Transformer for Semantic Segmentation

Lie et al. [75] proposed a novel vision Transformer called Swin Transformer, a hierarchical Transformer with shifted

windows used for general-purpose computer vision. To improve performance, offset windowing restricts self-attention computation to non-overlapping local windows while permitting cross-window connections. This hierarchical architecture can simulate various sizes and has a linear computational cost with image scalability. Swin Transformer used for image classification 86.4 accuracy on ImageNet-1K [76], dense prediction tasks including object identification 58.7 box AP on COCO, and semantic segmentation 53.5 mIoU on ADE20K [77].

Zheng et al. [78] introduced a sequence-to-sequence prediction framework for semantic segmentation. For the first time, authors have eliminated the need for FCN and solved a restricted receptive field problem, unlike current FCN-based techniques that use dilated convolutions and attention modules at the component level. This encoder may be coupled with a primary decoder to build a robust segmentation model called SEgmentation TRansformer (SETR). SETR uses a pure transformer (no convolution or resolution reduction) to encode an image as a patch sequence. MaX-DeepLab [79] is the first end-to-end model for panoptic segmentation that automatically infers masks and classes without the need for hand-coded priors such as object centres or boxes.

The Dense Prediction Transformer (DPT) [80] is a neural network design that successfully uses visual transformers for dense prediction problems. The monocular depth estimation and semantic segmentation tests demonstrate that the given architecture generates more fine-grained and globally coherent predictions than fully convolutional networks. As with previous work on transformers, when trained on large-scale datasets, the DPT reaches its full potential.

F. Transformer for Medical Image Segmentation

Segmentation of medical images is necessary for developing healthcare systems, particularly for disease diagnosis and treatment planning. The U-shaped architecture, commonly known as U-Net [81], achieved remarkable success in various medical image segmentation tasks. However, because convolution processes are intrinsically local, U-Net typically exhibits problems when representing long-range dependence clearly. To fully use the capabilities of Transformers, Chen et al. [82] presented TransUNet, which incorporates a fully global context by considering image features as sequences and effectively uses low-level CNN features via a U-shaped hybrid architectural design. Several experiments were conducted to evaluate the proposed TransUNet system and validate its performance in various scenarios, including 1) model scaling, 2) the number of skip-connections, 3) patch size and sequence length 4) input resolution. TransUNet outperforms many competing methods, including CNN-based self-attention methods, as an alternate framework to the current FCN-based systems for medical image segmentation. Fig. 6 shows the architecture of TransUNet.

Moreover, such a transformer model can also detect fraudulent activity in student identification documents. The model can scrutinize the identification document and highlight any discrepancies or irregularities. For example, the model can identify if the photo on the identification document has been manipulated digitally or if the document has been tampered with in any way. This fact not only streamlines the process of verifying student identities, making it quicker and more precise but also ensures the reliability of the verification process by detecting any fraudulent activity in student identification documents.

III. VISUAL TRANSFORMER VARIANTS

The development of the visual transformer has paved the way for significant advancements in computer vision. Since its inception, researchers have explored several variants of transformer architecture to further improve its performance on visual tasks. In this section, we discuss some of the notable variants of the visual transformer and their applications.

A. Deeper Visual Transformer

Zhou et al. found that in contrast to CNNs, the performance of vision transformers rapidly saturates as the number of convolutional layers increases. As the transformer progresses deeper, the attention maps become increasingly similar after a certain number of layers. The feature maps in the top layers of deep vision transformer models are often similar. It indicates that in the deeper layers of vision transformers, the self-attention mechanism cannot learn appropriate ideas for representation learning, preventing the model from achieving the predicted performance improvement. Zhou et al. [91] identified the problem of the vision transformers' attention collapsing as they progress deeper. They suggest a unique re-attention technique DeepViT to resolve it with the least amount of calculations and memory cost possible. Using Re-attention can sustain an improving performance when the depth of vision transformers increases.

The CaiT [92] network's operation involves two distinct processing phases. The first one, the self-attention stage, is similar to the vision transformer, except there is no class embedding. Second, a series of layers called the class-attention stage (CLS) compile the patch embeddings into a class embedding CLS, given to a linear classifier.

Wang et al. [93] propose a Pyramid vision transformer (PVT), a pure Transformer backbone suitable for dense prediction applications like semantic segmentation and object detection without convolutions. The authors create a progressive pyramid shrinking algorithm and a spatial-reduction attention layer for obtaining multi-scale feature maps

with minimal memory/computation resources. Extensive experimentation on semantic segmentation and object detection benchmarks demonstrates that PVT outperforms well-designed CNN when the parameters are equivalent. Fig. 7 compares the CNN architectures, the vision transformer, and the pyramid transform.

L. Yuan et al. [96] proposed a novel Token-to-Token Vision Transformer (T2T-ViT) model that can be trained entirely on ImageNet and attain performance equivalent to or better than CNN's. Using T2T-ViT, the image structure information is better modelled, and more features are provided. Thus T2T-ViT significantly exceeds the Vision Transformer features. It has a unique tokens-to-tokens (T2T) approach for tokenizing images incrementally and structurally aggregating tokens.

As a result of the improvements in computer vision and the enormous quantity of training data, many people feel Transformers are not appropriate for tiny datasets. The authors of this article [97] debunked the notion that transformers are data-hungry. The authors in [97] demonstrated that proposed Compact Convolution Transformers (CCT) can compete with state-of-the-art CNNs with appropriate data size and tokenization for the first time. Through a unique sequence pooling technique and convolutions, the suggested model eliminates the need for class tokens and positional embeddings.

Heo et al. [98] proposed a novel architecture called Pooling-based Vision Transformer (PiT) to use the pooling layers' advantages. The authors demonstrate that a commonly utilized design concept in CNN spatial dimensional transformation accomplished by pooling or convolution is ignored in transformer-based architectures, negatively affecting the model performance and the transformer architecture benefits from decreasing the spatial dimension. The authors initially examined ResNet and discovered that transforming it in terms of spatial dimension improves computing efficiency and generalization ability. To capitalize on the benefits of Vision Transformer, the authors proposed a PiT that integrates a pooling layer into Vision Transformer, and the PiT demonstrates that pooling layer benefits become effectively matched to Vision Transformer. As a result of considerably increasing the performance of the Vision Transformer architecture, the authors demonstrated that the pooling layer is critical for a self-attention-based design by considering the spatial interaction ratio. Moreover, extensive experiments showed that PiT outperforms the baseline on object detection, image classification, and robustness evaluation. Fig. 8 highlights the difference in dimensions of network architectures ResNet-50, Vision Transformer, and PiT.

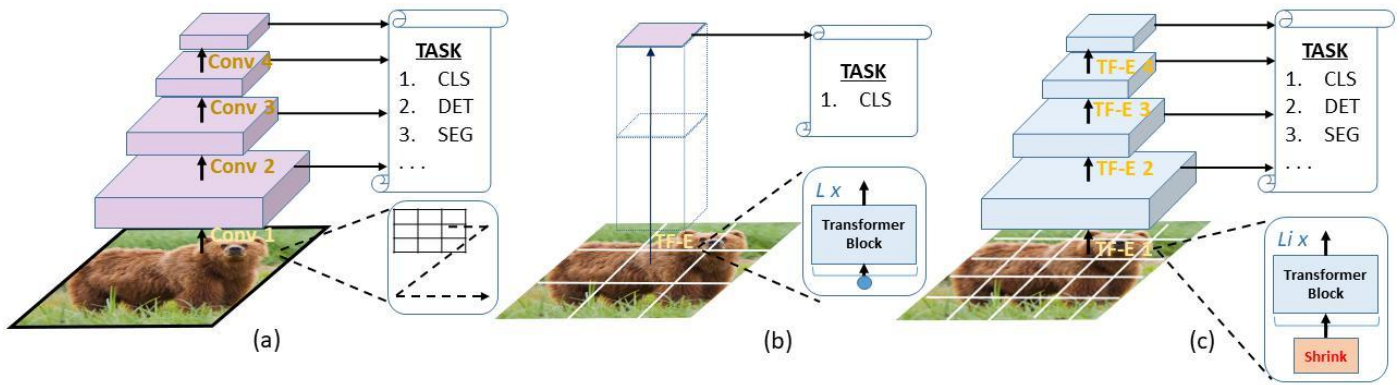


Fig. 7. (a) CNNs: ResNet [94], VGG [95], etc. (b) Vision transformer [58] (c) Pyramid vision transformer [93].

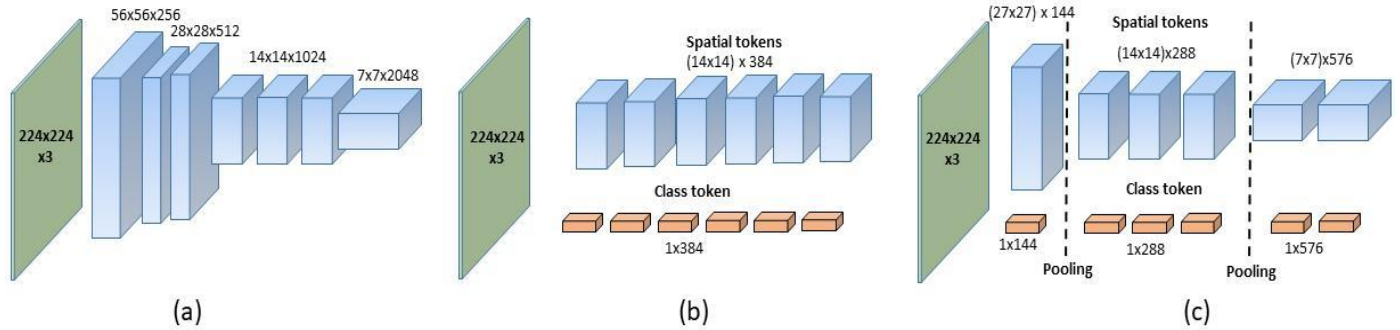


Fig. 8. A schematic diagram illustrates dimension variation in network architectures. (a) ResNet-50, (b) ViT-S/16, (c) PiT-S. image credit [98].

B. Hybrid Transformers

To improve image super-resolution, Z. Lu et al. [99] propose an *Efficient Super-Resolution Transformer (ESRT)*. ESRT is a hybrid Transformer that uses a CNN-based SR network to extract deep features. Backbones for the ESRT include the lightweight CNN (LCB) and lightweight Transformer (LTB). LCB is a low-cost SR network extracting deep SR features by dynamically changing the feature map's size. LTB consists of an efficient Transformer (ET) that consumes less GPU Memory space and benefits from the uniquely efficient multi-head attention (EMHA). The Reformer [100] is a Transformer model capable of processing context windows of up to 1 million words on a single accelerator with only 16GB of memory. Reformer combines two critical approaches for resolving the attention and memory allocation issues that limit the applicability of the transformer to lengthy context windows. The reformer uses locality-sensitive hashing (LSH) to decrease the complexity of attending to long sequences and reversible residual layers to maximize the usage of available memory.

H. Wu et al. [101] introduced a novel architecture, the *Convolutional Vision Transformer (CvT)* that enhances the performance and efficiency of the Vision Transformer by incorporating convolutions into the Vision Transformer. The author's findings indicate that positional encoding, a critical component of existing Vision Transformers, can be safely omitted from the CvT model, simplifying the design for higher-resolution vision applications.

Chu et al. [102] proposed a highly efficient and direct implementation of two architectures - *Twins-PCPVT* and

Twins-SVT vision transformer designs. *Twins-PCPT* is based on PVT [93] and CPVT [103] and utilizes global attention. *Twins-SVT* is based on the proposed SSSA, consisting of two distinct attention operations: locally-grouped self-attention (LSA) and globally subsampled attention (GSA). Both transformer models established new benchmarks for image classification, semantic/instance segmentation, and object detection.

Zhang et al. introduced the *Nested Transformer (NesT)* [104]. The block aggregation function is essential for enabling non-local information transmission across blocks. The accuracy of a NesT trained on ImageNet for 100/300 epochs is 82.3 per cent compared to other techniques [19], [105] that achieved up to 57% parameter reduction. A NesT with 6M parameters trained from scratch on CIFAR10 [106] achieves 96% accuracy using a single GPU.

Visual Transformers (VT) [107] defines the problem in the semantic token space, intending to represent and process high-level concepts in images using visual tokens. Moreover, different parts of the image have different meanings due to their different content. Note that this is entirely different from the transformer that processes information in pixel space (such as Vision Transformer, DeiT, IPT, etc.) because the amount of calculation differs by multiple orders of magnitude. The author [107] uses the spatial attention mechanism to convert the feature map into compact semantic tokens. Then input these tokens into a Transformer, and use the unique functions of the transformer to capture the connection between the tokens. In this way, VT can 1) Focus on those relatively important areas instead of treating all pixels equally like CNN. 2) Encode

semantic concepts in visual tokens instead of modelling all concepts in all images. 3) Use the Transformer to model the relationship between tokens. The VT model is used in classification tasks (Model Base: ResNet, Dataset: ImageNet, reduced by 6.9 times, increased by 4.6- 7 Accuracy) and semantic segmentation tasks (Model Base: FPN, Dataset: LIP and COCO-stuff reduced by 6.4 times the amount of calculation, the increase point 0.35 mIoU) has achieved excellent performance.

C. Supervised Learning in Vision Transformer

Supervised learning enables the transformer to learn a bottleneck representation in which the content and context are mixed around the class token. This results in a relatively superficial data model, and its association with labels needs many training examples. On the other hand, unsupervised learning uses the information redundancy and complementarity inherent in image data by learning to rebuild local content through context integration [108].

In self-supervised learning, no concept whatsoever of labelled data for the training. Self-supervised techniques can be classified broadly as generative or discriminative [109]. Generative methods learn to predict the data distribution. However, data modelling is inherently computationally expensive and may not be required in all cases for representation learning. Discriminative methods, generally implemented in a contrastive learning framework [110] or through pretext tasks [111], have the capacity to create more generalized representations with minimal computing needs.

Auto et al. [112] proposed a *Self-supervised vision Transformer (SiT)*, a unique approach for learning visual representations without supervision. Using the autoencoder transformer's inherent capacity to perform multitask learning, it created a robust self-supervised system that optimizes reconstruction, rotation classification, and contrastive losses concurrently. The last utilizes the strength of the transformer to train SiT to perform three distinct tasks: image reconstruction, rotation prediction, and contrastive learning.

Bao et al. [113] introduced a self-supervised vision representation model *BEiT*, which stands for Bidirectional Encoder representation from image Transformers. The authors proposed a masked image modelling task to pre-train vision Transformers in a self-supervised manner. In pre-training, each image contains two perspectives - image patches and visual tokens. First, "tokenize" the original image into visual tokens. Then, using a random masking technique, feed specific image patches into the backbone Transformer. The purpose of the pre-training is to reconstruct the original visual tokens from the damaged image patches. After pre-training BEiT, fine-tune model parameters directly on downstream tasks by superimposing task layers on the pre-trained encoder. Experiments on image classification and semantic segmentation demonstrate that our model outperforms prior pre-training approaches. For example, base-size BEiT achieves 83.2% top-1 accuracy on ImageNet-1K with the same

configuration, considerably surpassing DeiT training from scratch at 81.8% [19].

D. Video Transformer

Following the recent success of vision transformer models in image classification, Arnab et al. [114] presented pure-transformer-based video classification models *Video Vision Transformer (ViViT)*. To efficiently handle a high count of Spatiotemporal tokens, the authors in [114] constructed multiple model variations that factorize the transformer encoder's many components across spatial and temporal dimensions. The authors in [113] demonstrated how to use additional regularisation and pre-trained models to compensate for the fact that video datasets are often smaller than the image datasets on which Vision Transformer was trained.

The *VisTR*, a new video instance segmentation framework based on Transformers, considers the video in-stance segmentation (VIS) problem an end-to-end concurrent sequence decoding/prediction issue [115]. Fig. 9 shows the architecture of VisTR. The paradigm is qualitatively distinct from previous techniques, streamlining the whole process significantly. VisTR approaches the VIS problem from a novel similarity-based perspective. Segmentation was used to determine pixel-level similarity, whereas tracking was used to determine instance-to-instance similarity. Thus, tracking instances occurs naturally and smoothly in the instance segmentation context. VisTR's success is developing a novel technique, such as sequence matching and segmentation, optimized for the framework. This well-designed method enables monitoring and segmenting instances at the sequence level in their entirety. ViSTR is composed of four major components: 1) a CNN backbone that extracts feature representations from multiple images, 2) an encoder-decoder Transformer that models the relationships between pixel-level and instance-level features and decodes them, 3) an instance sequence matching module that supervises the model, and 4) an instance sequence segmentation module that outputs the final mask sequences. VisTR outperforms other techniques that employ a single model on the YouTube-VIS dataset, reaching 40.1% in mask mAP at 57.7 frames per second.

Fan et al. [116] introduce *Multi-scale Vision Transformers (MViT)* for video and image recognition by fusing the foundational concept of multi-scale feature hierarchies with transformer models. Multi-scale Transformers feature many scale stages with varying degrees of channel resolution.

The industry's high demand for autonomous driving has led to a surge of interest in three-dimensional object detection, resulting in several practical three-dimensional object detection algorithms [117]. Yuan et al. [55] proposed a *Temporal-Channel transformer* to represent spatial-temporal and channel domain relationships for video object detection from Lidar data. The transformer's unique architecture encodes temporal-channel information for many frames, whereas the decoder decodes spatial-channel information for the current frame voxel-by-voxel.

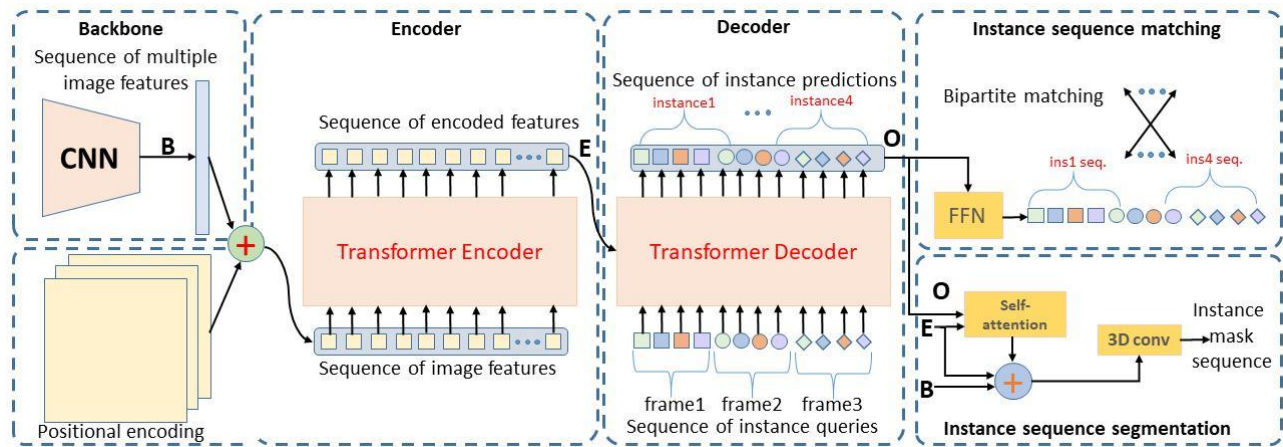


Fig. 9. The architecture of VisTR. image credit ([115]).

IV. MULTIMODAL TASK

With the increasing demand for models that can process both visual and textual inputs, the application of the visual transformer has been extended to multimodal tasks. In this section, we explore the capabilities of the visual transformer in handling multimodal inputs and review some of the recent developments in this area.

Recent breakthroughs in deep learning have resulted in significant advancements in computer vision and natural language processing. These accomplishments enable the integration of vision and language and multimodal learning tasks such as image captioning [118-119], image-text matching, visual grounding [120], and visual question answering [121]. Yu et al. present a novel framework for picture captioning called the Multimodal Transformer (MT) [122]. The MT comprises an image encoder that creates visual representations using deep self-attention learning and a caption decoder that converts the encoder’s visual characteristics to textual captions. Liu et al. [123] explore image captioning as a sequence-to-sequence prediction problem. Li et al. proposed CaPTion Trans- formerR (CPTR), a complete Transformer model, to replace the usual “CNN+Transformer” approach. CPTR is convolution-free and can model global context information at each encoder layer. Evaluation results on the famous MS COCO [68] dataset indicate that the CPTR technique is more successful than “CNN+Transformer” networks. Detailed visualizations illustrate that the CPTR model can use long-range dependencies from the start and that the decoder’s “words-to-patches” attention can pay close attention. The Conditional Position encodings Visual Transformer (CPVT) [103] sub-statutes the predefined positional embeddings used in Vision Transformer with conditional position encodings (CPE), allowing transformers to analyze input images of any size without interpolation.

Hu and Singh [124] developed a *Unified Transformer (UniT)* encoder-decoder model that accepts pictures and(or) text as input and trains on various tasks ranging from visual perception and language comprehension to combined vision-language reasoning. UniT consists of encoding modules that encode each input modality as a sequence of hidden states, a transformer decoder over the encoded input modalities, and

task-specific output heads that apply task-specific output heads to the decoder hidden states to generate the final predictions for each task. Desai and Johnson [125] proposed that visual representations from textual annotations (VirTex) are a pre-training technique for visual representations that use semantically dense captions. First, VirTex jointly trains CNN and Transformer to create natural language captions for images from scratch. Then, apply the newly acquired characteristics to subsequent visual recognition tasks. A Decision Transformer [126] architecture encodes states, actions, and returns using modality-specific linear embeddings and a positional episodic time step encoding. Fig. 10 shows the architecture of the Decision transformer. Tokens are fed into a GPT architecture, which uses a causal self-attention mask to predict behaviors auto-regressively.

The vision transformer architecture is integrated into generative adversarial networks (GANs) for image generation. Regularisation methods for GANs that are now available do not interact well with self-attention, resulting in significant training instability. GANs using Vision Transformers are trained using novel regularization techniques. ViTGAN beats the existing CNN-based StyleGAN2 method on the CIFAR-10, CelebA [127], and LSUN bedroom datasets [128].

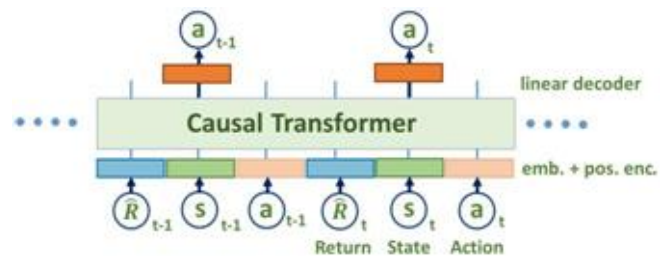


Fig. 10. Decision transformer architecture. Image credit [126].

V. FUTURE RESEARCH DIRECTIONS OF VISUAL TRANSFORMERS

As the field of Visual Transformers continues to develop, numerous potential avenues for future research exist. This section focuses on the summary of lessons learned from the discussions made throughout the article and future research directions. These directions offer valuable opportunities to enhance the performance and capabilities of Vision Transformers.

A. Lessons Learnt

1) *Adaptation of vision transformers:* The survey highlights the adaptability of Vision Transformers in their use for image processing, which was previously known in natural language processing. This fact demonstrates the chance to use already-developed skills and methods from one field to another, expanding the possible uses of Vision Transformers.

2) *Image enhancement and classification:* The survey shows how well Vision Transformers perform various image enhancement tasks, such as lowering noise, raising contrast, and boosting resolution. Additionally, their efficient use in image classification tasks demonstrates their capacity to extract significant representations from images and achieve competitive performance levels.

3) *State-of-the-Art object detection:* The study demonstrates the remarkable object detection performance of Vision Transformers, which outperforms conventional approaches and produces cutting-edge results. This fact shows the main contribution that Vision Transformers make to the field by enhancing robustness and accuracy in object localization and identification key processes.

4) *Automation in education systems:* The survey looks at how student verification processes used in educational institutions utilize Vision Transformers. Vision Transformers can streamline administrative operations, increasing speed and accuracy by automating the identity of the verification process. This fact demonstrates the crucial role that Vision Transformers have played in changing and strengthening the administrative procedures used in educational institutions.

5) *Multimodal processing:* The survey explores Vision Transformers' ability to handle multimodal tasks that require the processing and fusion of data from many modalities, including text, picture, and audio. This fact demonstrates the potential of Vision Transformers to enable thorough comprehension and analysis of complex data, leading to breakthroughs in areas like cross-modal retrieval, multimodal sentiment analysis, and visual question answering. The review demonstrates the significant contributions and developments made by Vision Transformers for image processing through an in-depth examination. The main findings from this overview highlight the flexibility of Vision Transformers, their effectiveness in image enhancement and classification, their cutting-edge performance in object detection, their potential to automate administrative tasks in educational systems, and their proficiency in handling multimodal data. These priceless insights provide direction for future study and practical use of Vision Transformers across numerous areas, encouraging further development in computer vision.

B. Future Research Directions

1) *Efficiency optimization:* Recent research has focused a lot of attention on finding ways to increase the effectiveness of Visual Transformers. There is a rising need to create methods to increase Visual Transformers' efficiency without sacrificing performance due to the demand for real-time and resource-

constrained applications. Recent studies have looked into several solutions to this problem. For instance, researchers have studied techniques for sparse attention, which concentrate on focusing just on relevant areas of input to lighten the total computing load. Also, low-rank approximations, which approximate the attention matrices with low-rank structures and save a significant amount of computation, have been studied. Additionally, to decrease the memory footprint and inference time of Visual Transformers, researchers have looked into model reduction techniques like pruning or quantization. Recent developments in Visual Transformer efficiency show tremendous promise for enabling implementation in resource-constrained contexts while preserving their efficacy and performance.

2) *Robustness and generalization:* A significant area of research continues to be strengthening the robustness and generalization abilities of Visual Transformers, with current developments tackling the difficulties presented by real-world settings. Researchers have been looking into cutting-edge methods to lessen the effects of occlusions, which frequently impair accurate object detection in complex surroundings, in the pursuit of enhanced resilience. Recent research has looked at techniques to improve performance in obstructed situations, such as partial occlusion handling through attention processes or occlusion-aware training procedures. Additionally, the danger of adversarial attacks has been elevated to a top priority when using computer vision models. By using competitive training techniques or defence mechanisms against such attacks, researchers have made substantial progress toward creating robust Visual Transformers that can tolerate adversarial perturbations. In addition, recent research initiatives have focused on addressing domain transitions. It has been investigated to enhance the generalization abilities of Visual Transformers across various datasets or real-world domains using techniques like domain adaption or domain generalization. Researchers are making progress towards giving Visual Transformers the robustness and generalization skills they need to meet the challenges posed by many complicated real-world circumstances by actively taking into account these recent developments.

3) *Interpretability and explainability:* Various computer vision challenges have revealed impressive performance from Visual Transformers. However, it is difficult to comprehend the logic behind their forecasts because of their lack of interpretability. The recent research aims to overcome this drawback by investigating ways to make Visual Transformers easier to understand. A possible strategy is using attention visualization methods to draw attention to the areas of an image impacting judgment. Researchers and users can learn more about the particular characteristics or regions that the Visual Transformer concentrates on while making predictions by visualizing the attention maps. In addition, techniques like gradient-based attribution approaches and saliency maps have been used to pinpoint the most crucial input features influencing the result. These methods aid in identifying the

primary determinants of the Visual Transformer's choice, enhancing the predictability and interpretability of results. Future research aims to provide users with more transparent and interpretable Visual Transformers, enabling improved understanding and utilization of these potent models in practical applications. The last will be accomplished by continuing to investigate and improve these methodologies.

4) *Multi-modal and cross-modal learning*: A fascinating research area that will help us interpret complicated visual data better is the extension of Visual Transformers to handle multimodal and cross-modal data. Inquiries into integrating Visual Transformers with various modalities, such as text, audio, and depth information, have advanced significantly in recent years. For instance, using the strength of Visual Transformers to interpret visual data alongside textual context, researchers have created unique architectures that integrate vision and language models. Tasks such as image captioning, visual question answering, and cross-modal retrieval have all benefited from this integration. Additionally, research into using audio data in Visual Transformers has produced promising outcomes for tasks like sound event recognition or audio-visual scene analysis. Another fascinating development is integrating depth information with Visual Transformers, which enables comprehensive scene interpretation and 3D perception. Visual Transformers can deliver a thorough and holistic comprehension of complicated visual data by successfully integrating and learning from several modalities, pushing the limits of computer vision research and applications. Continued study in this field can reveal fresh perspectives and enhance Visual Transformers' capacity to handle multimodal and cross-modal input.

5) *Incremental and continual learning*: Recent research has focused heavily on enabling Visual Transformers to continually learn from streaming or updating data since it allows models to adapt to changing contexts and evolving concepts. The flexibility and adaptability of Visual Transformer may be enhanced by recent developments in incremental learning methods. Rehearsal approaches, which save and playback a portion of previously observed data during training to reduce catastrophic forgetting, are one noteworthy strategy. Research has also looked into methods like lifelong learning, where the model gradually picks up new skills while holding on to knowledge from earlier jobs. As a result, Visual Transformers can continuously improve their skills without compromising how well they do previously mastered jobs. Strategies like adaptive learning rates, dynamic network designs, and online learning algorithms have been investigated to address the problem. Visual Transformers can effectively learn from evolving data streams, improve their performance, and keep current knowledge by concentrating on incremental learning and devising ways to adapt to new classes or concepts over time. More research is needed in this field to make Visual Transformers more flexible and adaptable in practical applications and dynamic contexts.

6) *Attention mechanism exploration*: Research on Visual Transformers in recent years has concentrated on understanding and improving attention mechanisms to improve their effectiveness. Different attention types that can improve the modelling skills of Visual Transformers are the subject of one area of research. For instance, non-local attention mechanisms have drawn attention to their ability to identify distant relationships in pictures or movies, facilitating a better comprehension of the whole context. Another interesting approach is sparse attention, which tries to keep good performance while reducing computing complexity by focusing only on pertinent areas or pixels inside an input. Additionally, researchers have looked at the usage of learned attention masks, in which attention weights are dynamically computed based on the input data, enabling the model to assign adaptively attention to the most informative regions. The performance and modelling skills of Visual Transformers could be significantly improved by these latest developments in attention mechanisms. Researchers can open new doors for developing computer vision and expanding the capabilities of Visual Transformers in various applications by exploring these attention variants and continuing to innovate in this area.

7) *Domain-specific adaptation*: Various computer vision challenges have revealed impressive performance from Visual Transformers. However, because of the particular traits and demands of such areas, its application to specific tasks or domains frequently presents difficulties. Future research efforts can concentrate on investigating domain-specific adaptation methods to modify Visual Transformers for specific application domains. Recent studies have begun to explore domain adaptation techniques that use labelled data from the target domain to align the model's representation with the domain-specific features. To adapt Visual Transformers for tasks like disease diagnosis, organ segmentation, or anomaly detection, for instance, researchers in the field of medical imaging have investigated strategies like transfer learning or fine-tuning on medical datasets. Although Visual Transformers have demonstrated potential in satellite image analysis, more study is required to create domain-specific adaptations to deal with issues like size variation, heterogeneous data sources, or a lack of labelled data. In a similar way, in robotics, Visual Transformers can be configured to perform visual perception tasks in specific robotic applications, such as robot localization, object recognition, and scene interpretation. Researchers can bridge the gap between Visual Transformers and certain application areas, enabling higher performance and overcoming the particular difficulties encountered in those domains by concentrating on domain-specific adaption strategies. The investigation of these methods holds promise for releasing Visual Transformers' full potential across many specialized fields and advancing computer vision in particular application areas.

8) *Data-efficient learning*: Visual Transformers have displayed outstanding performance in computer vision

applications, although their training frequently necessitates large amounts of labelled data. Recent research has concentrated on investigating data-efficient learning approaches to lessen the dependence on sizable annotated datasets and enable efficient learning with few labelled examples. In this regard, semi-supervised learning strategies have drawn interest since they use labelled and unlabeled data during training. Visual Transformers can gain from a larger training set and perform better by using the quantity of unlabeled data and incorporating it into the learning process. Another exploratory route, which seeks to learn representations solely from unlabeled data, is unsupervised learning. These techniques allow models to develop helpful presentations from unannotated data that may be applied to subsequent tasks. Unsupervised learning has recently made significant strides in several computer vision areas, including picture categorization, object recognition, and image synthesis. Researchers can harness the potential of Visual Transformers in situations with little labelled data by exploring data-efficient learning techniques, making it possible to deploy these models more frugally and widely in various applications.

V. CONCLUSION

This article discusses critical self-attention architectures and examines in detail transformer models for various image-processing applications. We comprehensively discuss the strengths and weaknesses of existing techniques, particularly the possible future research directions. With a particular emphasis on general image processing problems, this survey offers a unique perspective on recent advances in self-attention and Transformer-based techniques. We discuss state-of-the-art self-attention models for semantic and instance segmentation, image classification, object detection, image captioning, video analysis and classification, multi-model tasks, and three-dimensional data analysis. We hope our work will spark interest among the image-processing community in maximizing the applications of vision-transformed models. Transformer models are pretty complicated from the perspective of parameters, computing time, and resources required. Visualizing and comprehending essential parts in an image for classification purposes is still a problem in transformers, and spatially accurate activation-specific representations are necessary [129]. The use of a vision transformer model in university education systems facilitating the detection of fraudulent activities in student identification documents is also highlighted. The model can thoroughly examine the identification document, detecting inconsistencies or anomalies as highlighting fraudulent activity.

ACKNOWLEDGMENT

This paper is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01. The paper reflects only the author's view and the Agency is not responsible for any use that may be made of the information it contains.

REFERENCES

- [1] Y. LeCun, "Backpropagation applied to digit recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [2] I. Sutskever, O. Vinyals, and Q. Le, "Sequence to sequence learning with neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 4, no. January, pp. 3104–3112, 2014.
- [3] P. Velicković, A. Casanova, P. Lio, G. Cucurull, A. Romero, and Y. Bengio, "Graph attention networks," in 6th Int. Conf. Learn. Represent. ICLR 2018 - Conf. Track Proc, 2018, pp. 1–12.
- [4] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, "Spatial transformer networks," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 2017–2025, 2015.
- [5] K. Xu, M. Zhang, J. Li, S. Du, K. Kawarabayashi, and S. Jegelka, "How neural networks extrapolate: From feedforward to graph neural networks," 2020, available: [Online]. Available: <http://arxiv.org/abs/2009.11848>.
- [6] P. Ramachandran, I. Bello, N. Parmar, A. Levskaya, A. Vaswani, and J. Shlens, "Stand-alone self-attention in vision models," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
- [7] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. m, no. Nips, p. 5999–6009, 2017.
- [8] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf.*, vol. 1, no. Mlm, pp. 4171–4186, 2019.
- [9] M. Peters, "Knowledge enhanced contextual word representations," *emnlp-ijcnlp 2019 - 2019 conf.*, in *Empir. Methods Nat. Lang. Process.* 9th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf, 2020, pp. 43–54.
- [10] C. Raffel, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, pp. 1–67, 2020.
- [11] S. Mehta, M. Ghazvininejad, S. Iyer, L. Zettlemoyer, and H. Hajishirzi, "Delight: Deep and light-weight transformer," 2020, available: [Online]. Available: <http://arxiv.org/abs/2008.00623>.
- [12] Z. Fan, "Mask attention networks: Rethinking and strengthen transformer," pp. 1692–1701, 2021.
- [13] M. Shoybi, M. Patwary, R. Puri, P. LeGresley, J. Casper, and Catanzaro, "Megatron-lm: Training multi-billion parameter language models using model parallelism," 2019, available: [Online]. Available: <http://arxiv.org/abs/1909.08053>.
- [14] A. Roy, M. Saffar, A. Vaswani, and D. Grangier, "Efficient content-based sparse attention with routing transformers," *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 53–68, 2021.
- [15] Z. Dai, Z. Yang, Y. Yang, J. Carbonell, Q. Le, and R. Salakhutdinov, "Transformer-xl: Attentive language models beyond a fixed-length context," in *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, 2020, pp. 2978–2988.
- [16] H. Yan, B. Deng, X. Li, and X. Qiu, "Tener: Adapting transformer encoder for named entity recognition," 2019, available: [Online]. Available: <http://arxiv.org/abs/1911.04474>.
- [17] X. Li, H. Yan, X. Qiu, and X. Huang, "Flat: Chinese ner using flat-lattice transformer," pp. 6836–6842, 2020.
- [18] R. Girdhar, J. Carreira, C. Doersch, and A. Zisserman, "Video action transformer network," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, 2019, pp. 244–253.
- [19] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and Jegou, "Training data-efficient image transformers distillation through attention," pp. 1–22, 2020, available: [Online]. Available: <http://arxiv.org/abs/2012.12877>.
- [20] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5790–5799.
- [21] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and Zagoruyko, "End-to-end object detection with transformers," *LNCS*, vol. 12346, pp. 213–229, 2020.

- [22] Y. Wang, "Vistr: End-to-end instance segmentation with transformers," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2105.00637>.
- [23] A. Ramesh, "Zero-shot text-to-image generation," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2102.12092>.
- [24] W. Su, "Vi-bert: Pre-training of generic visual-linguistic representations," pp. 1–16., 2019, available:. [Online]. Available: <http://arxiv.org/abs/1908.08530>.
- [25] Y. Xu, H. Wei, M. Lin, Y. Deng, K. Sheng, M. Zhang, F. Tang, Dong, F. Huang, and C. Xu, "Transformers in computational visual media: A survey," *Computational Visual Media*, vol. 8, pp. 33–62, 2022.
- [26] S. Jamil, M. Jalil Piran, and O.-J. Kwon, "A comprehensive survey of transformers for computer vision," *Drones*, vol. 7, no. 5, p. 287, 2023.
- [27] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, Y. Zhang, Shi, J. Fan, and Z. He, "A survey of visual transformers," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [28] J. Selva, A. S. Johansen, S. Escalera, K. Nasrollahi, T. B. Moeslund, and A. Clapes, "Video transformers: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [29] S. Khan, M. Naseer, M. Hayat, S. Zamir, F. Khan, and M. Shah, "Transformers in vision: A survey," pp. 1–28., 2021, available:. [Online]. Available: <http://arxiv.org/abs/2101.01169>.
- [30] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," pp. 53 040–53 065., 2019.
- [31] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2106.04554>.
- [32] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, Xiao, C. Xu, Y. Xu et al., "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.
- [33] C.-F. R. Chen, Q. Fan, and R. Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 357–366.
- [34] M. Gehrig and D. Scaramuzza, "Recurrent vision transformers for object detection with event cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 884–13 893.
- [35] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling vision transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 104–12 113.
- [36] Y. Yang, L. Jiao, X. Liu, F. Liu, S. Yang, Z. Feng, and X. Tang, "Transformers meet visual learning understanding: A comprehensive review," *arXiv preprint arXiv:2203.12944*, 2022.
- [37] J. Guo, K. Han, H. Wu, Y. Tang, X. Chen, Y. Wang, and C. Xu, "Cmt: Convolutional neural networks meet vision transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 175–12 185.
- [38] K. He, C. Gan, Z. Li, I. Rekić, Z. Yin, W. Ji, Y. Gao, Q. Wang, Zhang, and D. Shen, "Transformers in medical image analysis: A review," *Intelligent Medicine*, 2022.
- [39] A. A. Aleissae, A. Kumar, R. M. Anwer, S. Khan, H. Cholakkal, G.-Xia, and F. S. Khan, "Transformers in remote sensing: A survey," *Remote Sensing*, vol. 15, no. 7, p. 1860, 2023.
- [40] "Vision transformer explained — papers with code," accessed Oct. 11, 2021). [Online]. Available: <https://paperswithcode.com/method/vision-transformer>
- [41] J. Thickstun, "The transformer model equations," pp. 1–5., 2019, available:. [Online]. Available: <https://homes.cs.washington.edu/thickstn/docs/transformers.pdf>.
- [42] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *3rd International Conference on Learning Representations, ICLR 2015*, 2015, p. 1–15.
- [43] M. Luong, H. Pham, and C. Manning, "Effective approaches to attention-based neural machine translation," in *Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process*, 2015, pp. 1412–1421..
- [44] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, 2018, pp. 7794–7803..
- [45] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus," pp. 1–9., 2016, [Online]. Available: <http://arxiv.org/abs/1606.08415>.
- [46] J. Ba, J. Kiros, and G. Hinton, "Layer normalization," 2016, available:. [Online]. Available: <http://arxiv.org/abs/1607.06450>.
- [47] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.
- [48] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019, available:. [Online]. Available: <http://arxiv.org/abs/2007.07582>.
- [49] T. Brown, "Language models are few-shot learners," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, 2020.
- [50] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, pp. 1–18., 2019.
- [51] Y. Liu, "Roberta: A robustly optimized bert pretraining approach," 2019, available:. [Online]. Available: <http://arxiv.org/abs/1907.11692>.
- [52] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and Soricut, "Albert: A lite bert for self-supervised learning of language representations," pp. 1–17., 2019, available:. [Online]. Available: <http://arxiv.org/abs/1909.11942>.
- [53] C. Rosset, "Turing-nlg: A 17-billion-parameter language model by microsoft," *Microsoft Research*, p. 17, 2020, accessed Sep. 05, 2021). [Online]. Available: <https://www.microsoft.com/en-us/research/blog/turing-nlg-a>
- [54] H. Xu, "Pre-trained models: Past, present and future," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2106.07139>.
- [55] Z. Chen, L. Xie, J. Niu, X. Liu, L. Wei, and Q. Tian, "Visformer: The vision-friendly transformer," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2104.12533>.
- [56] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and Vaswani, "Bottleneck transformers for visual recognition," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2101.11605>.
- [57] H. Zhao, L. Jiang, J. Jia, P. Torr, and V. Koltun, "Point transformer," 2020, available:. [Online]. Available: <http://arxiv.org/abs/2012.09164>.
- [58] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, available:. [Online]. Available: <http://arxiv.org/abs/2010.11929>.
- [59] M. Chen, "Generative pretraining from pixels," in *37th Int. Conf. Mach. Learn. ICML 2020*, 2020, vol. PartF16814, pp. 1669–1681..
- [60] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Weinberger, "Deep networks with stochastic depth," *LNCS*, vol. 9908, pp. 646–661., 2016.
- [61] E. Hoffer, T. Ben-Nun, I. Hubara, N. Giladi, T. Hoefler, and D. Soudry, "Augment your batch: better training with larger batches," 2019, available:. [Online]. Available: <http://arxiv.org/abs/1901.09335>.
- [62] B. Graham, "Levit: a vision transformer in convnet's clothing for faster inference," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2104.01136>.
- [63] C.-F. Chen, Q. Fan, and R. Panda, "Crossvit: Cross-attention multi-scale vision transformer for image classification," 2021, available:. [Online]. Available: <http://arxiv.org/abs/2103.14899>.
- [64] K. Han, A. Xiao, E. Wu, J. Guo, C. Xu, and Y. Wang, "Transformer in transformer," pp. 1–12., 2021, available:. [Online]. Available: <http://arxiv.org/abs/2103.00112>.
- [65] H. Chen, "Pre-trained image processing transformer," 2020, available:. [Online]. Available: <http://arxiv.org/abs/2012.00364>.
- [66] A. Vaswani, "Tensor2tensor for neural machine translation," in *AMTA 2018 - 13th Conference of the Association for Machine Translation in the Americas, Proceedings*, 2018, vol. 1, p. 193–199.
- [67] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," pp. 1–16., 2020, available:. [Online]. Available: <http://arxiv.org/abs/2010.04159>.

- [68] T. Lin, "Microsoft coco: Common objects in context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*, no. PART 5, pp. 740–755., 2014.
- [69] M. Zheng, P. Gao, X. Wang, H. Li, and H. Dong, "End-to-end object detection with adaptive clustering transformer," 2020, available: [Online]. Available: <http://arxiv.org/abs/2011.09315>.
- [70] Z. Dai, B. Cai, Y. Lin, and J. Chen, "Up-detr: Unsupervised pre-training for object detection with transformers," 2020, available: [Online]. Available: <http://arxiv.org/abs/2011.09094>.
- [71] Z. Sun, S. Cao, Y. Yang, and K. Kitani, "Rethinking transformer-based set prediction for object detection," 2020, available: [Online]. Available: <http://arxiv.org/abs/2011.10881>.
- [72] J. Beal, E. Kim, E. Tzeng, D. Park, A. Zhai, and D. Kislyuk, "Toward transformer-based object detection," 2020, available: [Online]. Available: <http://arxiv.org/abs/2012.09958>.
- [73] J. Yang, "Focal self-attention for local-global interactions in vision transformers," pp. 1–21., 2021, available: [Online]. Available: <http://arxiv.org/abs/2107.00641>.
- [74] S. He, H. Luo, P. Wang, F. Wang, H. Li, and W. Jiang, "Transreid: Transformer-based object re-identification," 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.04378>.
- [75] Z. Liu, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.14030>.
- [76] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," pp. 248–255., 2014-05.
- [77] B. Zhou, "Semantic understanding of scenes through the ade20k dataset," *Int. J. Comput. Vis.*, vol. 127, no. 3, pp. 302–321., 2019.
- [78] S. Zheng, "Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers," pp. 6881–6890., 2020, available: [Online]. Available: <http://arxiv.org/abs/2012.15840>.
- [79] H. Wang, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen, "Max-deeplab: End-to-end panoptic segmentation with mask transformers," pp. 5463–5474., 2020, available: [Online]. Available: <http://arxiv.org/abs/2012.00759>.
- [80] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.13413>.
- [81] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics*, vol. 9351, pp. 234–241., 2015.
- [82] J. Chen, "Transunet: Transformers make strong encoders for medical image segmentation," pp. 1–13., 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.04306>.
- [83] B. Yun, Y. Wang, J. Chen, H. Wang, W. Shen, and Q. Li, "Spectr: Spectral transformer for hyperspectral pathology image segmentation," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.03604>.
- [84] J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," pp. 1–18., 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.10662>.
- [85] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L. Chen, "Axial-deeplab: Stand-alone axial-attention for panoptic segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12349, pp. 108–126.
- [86] M. Guo, J. Cai, Z. Liu, T. Mu, R. Martin, and S. Hu, "Pct: Point cloud transformer," *Comput. Vis. Media*, vol. 7, no. 2, pp. 187–199., 2021.
- [87] N. Parmar, "Image transformer," in 35th Int. Conf. Mach. Learn. ICML 2018, 2018, vol. 9, pp. 6453–6462.,
- [88] Y. Jiang, S. Chang, and Z. Wang, "Transgan: Two pure transformers can make one strong gan, and that can scale up," 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.07074>.
- [89] K. Lee, H. Chang, L. Jiang, H. Zhang, Z. Tu, and C. Liu, "Vitgan: Training gans with vision transformers," pp. 1–13., 2021, available: [Online]. Available: <http://arxiv.org/abs/2107.04589>.
- [90] Z. Huang, J.-X. Du, and H.-B. Zhang, "A multi-stage vision transformer for fine-grained image classification," in 2021 11th International Conference on Information Technology in Medicine and Education (ITME). IEEE, 2021, pp. 191–195.
- [91] D. Zhou, "Deepvit: Towards deeper vision transformer," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.11886>.
- [92] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jegou, "Going deeper with image transformers," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.17239>.
- [93] W. Wang, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.12122>.
- [94] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, 2016, pp. 770–778..
- [95] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in 3rd International Conference on Learning Representations, ICLR 2015, 2015, p. 1–14.
- [96] L. Yuan, "Tokens-to-token vit: Training vision transformers from scratch on imagenet," 2021, available: [Online]. Available: <http://arxiv.org/abs/2101.11986>.
- [97] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, "Escaping the big data paradigm with compact transformers," 2021, available: [Online]. Available: <http://arxiv.org/abs/2104.05704>.
- [98] B. Heo, S. Yun, D. Han, S. Chun, J. Choe, and S. Oh, "Rethinking spatial dimensions of vision transformers," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.16302>.
- [99] Z. Lu, H. Liu, J. Li, and L. Zhang, "Efficient transformer for single image super resolution," pp. 1–13., 2021, available: [Online]. Available: <http://arxiv.org/abs/2108.11084>.
- [100] N. Kitaev, Kaiser, and A. Levskaya, "Reformer: The efficient transformer," pp. 1–12., 2020, available: [Online]. Available: <http://arxiv.org/abs/2001.04451>.
- [101] H. Wu, "Cvt: Introducing convolutions to vision transformers," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.15808>.
- [102] X. Chu, "Twins: Revisiting the design of spatial attention in vision transformers," pp. 1–14., 2021, available: [Online]. Available: <http://arxiv.org/abs/2104.13840>.
- [103] "Conditional positional encodings for vision transformers," 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.10882>.
- [104] Z. Zhang, H. Zhang, L. Zhao, T. Chen, and T. Pfister, "Aggregating nested transformers," pp. 1–18., 2021, available: [Online]. Available: <http://arxiv.org/abs/2105.12723>.
- [105] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in 36th Int. Conf. Mach. Learn. ICML 2019, 2019, vol. 2019-June, pp. 10 691–10 700.,
- [106] A. Krizhevsky, "Learning multiple layers of features from tiny images," 2009.
- [107] B. Wu, "Visual transformers: Token-based image representation and processing for computer vision," 2020, available: [Online]. Available: <http://arxiv.org/abs/2006.03677>.
- [108] M. Caron, "Emerging properties in self-supervised vision transformers," 2021, [Online]. Available: <http://arxiv.org/abs/2104.14294>.
- [109] M. Patacchiola and A. Storkey, "Self-supervised relational reasoning for representation learning," *Adv. Neural Inf. Process. Syst.*, vol. 2020-Decem, no. NeurIPS, 2020.
- [110] J. Donahue and K. Simonyan, "Large scale adversarial representation learning," *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, pp. 1–32., 2019.
- [111] S. Jenni and P. Favaro, "Self-supervised feature learning by learning to spot artifacts," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, 2018, pp. 2733–2742.,
- [112] S. Aitō, M. Awais, and J. Kittler, "Sit: Self-supervised vision transformer," pp. 1–10., 2021, available: [Online]. Available: <http://arxiv.org/abs/2104.03602>.

- [113] H. Bao, L. Dong, and F. Wei, "Beit: Bert pre-training of image transformers," pp. 1–16, 2021, available: [Online]. Available: <http://arxiv.org/abs/2106.08254>.
- [114] A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucic, and C. Schmid, "Vivit: A video vision transformer," 2021, available: [Online]. Available: <http://arxiv.org/abs/2103.15691>.
- [115] Y. Wang, "End-to-end video instance segmentation with transformers," p. 8741–8750, 2020.
- [116] H. Fan, "Multiscale vision transformers," 2021. [Online]. Available: <http://arxiv.org/abs/2104.11227>.
- [117] R. Liu, Z. Yuan, T. Liu, and Z. Xiong, "End-to-end lane shape prediction with transformers," 2020.
- [118] J. Lu, D. Batra, D. Parikh, and S. Lee, "Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks," *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 1–11, 2019.
- [119] J. Lu, V. Goswami, M. Rohrbach, D. Parikh, and S. Lee, "12-in-1: Multi-task vision and language representation learning," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*, 2020, pp. 10 434–10 443,.
- [120] H. Akbari, "Vatt: Transformers for multimodal self-supervised learning from raw video," *Audio and Text*, 2021, available: [Online]. Available: <http://arxiv.org/abs/2104.11178>.
- [121] L. Li, M. Yatskar, D. Yin, C.-J. Hsieh, and K.-W. Chang, "Visualbert: A simple and performant baseline for vision and language," pp. 1–14, 2019, available: [Online]. Available: <http://arxiv.org/abs/1908.03557>.
- [122] J. Yu, J. Li, Z. Yu, and Q. Huang, "Multimodal transformer with multi-view visual representation for image captioning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4467–4480, 2020.
- [123] W. Liu, S. Chen, L. Guo, X. Zhu, and J. Liu, "Cptr: Full transformer network for image captioning," pp. 1–5, 2021, available: [Online]. Available: <http://arxiv.org/abs/2101.10804>.
- [124] R. Hu and A. Singh, "Unit: Multimodal multitask learning with a unified transformer," 2021, available: [Online]. Available: <http://arxiv.org/abs/2102.10772>.
- [125] K. Desai and J. Johnson, "Virtex: Learning visual representations from textual annotations," 2020, available: [Online]. Available: <http://arxiv.org/abs/2006.06666>.
- [126] L. Chen, "Decision transformer: Reinforcement learning via sequence modeling," pp. 1–21, 2021, available: [Online]. Available: <http://arxiv.org/abs/2106.01345>.
- [127] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, 2015, pp. 3730–3738,.
- [128] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," 2015, available: [Online]. Available: <http://arxiv.org/abs/1506.03365>.
- [129] K. Han, "A survey on vision transformer," pp. 1–25, 2020, available: [Online]. Available: <http://arxiv.org/abs/2012.12556>.

Multimodal Contactless Architecture for Upper Limb Virtual Rehabilitation

Emilio Valdivia-Cisneros, Elizabeth Vidal*, Eveling Castro-Gutierrez
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

Abstract—The use of virtual rehabilitation systems for upper limbs has been implemented using different devices, and its efficiency as a complement to traditional therapies has been demonstrated. Multimodal systems are necessary for virtual rehabilitation systems since they allow multiple sources of information for both input and output so that the participant can have a personalized interaction. This work presents a simplified multimodal contactless architecture for virtual reality systems that focuses on upper limb rehabilitation. This research presents the following: 1) the proposed architecture 2) the implementation of a virtual reality system oriented to activities of daily living, and 3) an evaluation of the user experience and the kinematic results of the implementation. The results of the two experiments showed positive results regarding the implementation of a multimodal contactless virtual rehabilitation system based on the architecture. User experience evaluation showed positive values regard to six dimensions: perspicuity=2.068, attractiveness=1.987, stimulation=1.703, dependability=1.649, efficiency=1.517, and novelty=1.401. Kinematic evaluation was consistent with the score of the implemented game.

Keywords—Human computer interaction (HCI); multimodal; feedback; architecture; upper limb; rehabilitation; contactless

I. INTRODUCTION

Limited mobility is a prevalent dysfunction that is observed in patients suffering from neurological diseases such as Stroke, Epileptic Encephalopathy, Cerebral Palsy or Parkinson diseases [1, 2, 3, 4]. The importance of upper limb function rehabilitation is emphasized since upper limbs are used to manipulate objects and to interact physically in Activities of Daily Living (ADL) [5, 6].

Studies present different types of implementations for rehabilitation that respond to different kinds of patients' needs [7, 8, 9] however, sometimes medical conditions do not allow the use of wearable devices [10]. Contactless devices are an alternative for gesture recognition applications in healthcare [11] and permit the tracking of free and natural movements facilitating the user's mobility [12]. Important advances have been made in contactless approaches that grant safety and accuracy [13, 14].

Multimodal systems allow Virtual Reality Systems (VRS) to be implemented with multiple sources of information for both input and output so that the participant can have a personalized interaction with the system [15, 16]. Multimodal systems also consider multimodal feedback that consists of visual, auditory, and tactile feedback that can be combined to increase participant motivation and improve training effectiveness [17].

Literature shows different kinds of focus regarding the use of contactless devices, for example, the analysis of the contactless interactions of new users when they are learning and adapting [18]. Some other research focuses on the impact of a specific device in the rehabilitation process such as Leap Motion Controller [19, 20, 21] or Kinect [22, 23, 24, 25]. Some other studies refer to how virtual rehabilitation using contactless devices reinforce motivation [26, 27, 28, 29].

The main objectives of this work are: 1) to propose a multimodal contactless architecture for a virtual rehabilitation system for upper limbs; 2) implement a virtual rehabilitation system based on the proposed architecture; and 3) evaluate the acceptance of the VRS and the kinematic outcomes.

The rest of the paper is organized as follows: Section II presents the related works; Section III presents the architecture and experiments conducted. Section IV presents results and discussion, and finally Section V gives the conclusions.

II. RELATED WORK

Contactless devices for rehabilitation have been used for many years. Early research makes use of Kinect, and experiences have been carried out for various diseases such as stroke, cerebral palsy or Parkinson disease.

Huang [30] implemented recognition on arm movements. Therapists were able to adjust the rehabilitation movements based on the conditions of the participant. Pastor et al. [31] focused on increasing range of motion to improve functional use of the impaired upper extremity. They developed a game that requires patients to control a cursor on the screen by moving their hand.

Other research has focused on ADL, for example, the work in Adams [32] implemented activities for preparing meals with an avatar for recovery of upper extremities combining a virtual world and a Kinect™ sensor.

There is other research that has focused on active movements of the upper and lower limbs using a Kinect-based game system in addition to conventional therapy. The results showed that Kinect may have supplemental benefits for patients [33] [34].

In recent years, other contactless devices have become relevant in the virtual rehabilitation process due to their small size, accuracy, and low cost. Taraki et al. [35] presented the use of the Leap Motion Controller (LMC) for upper extremity rehabilitation to improve the joint range of motion, muscle strength, coordination, and fine motor functions of the hand and wrist in patients. The results showed quantitatively that

LMC should be used as an effective alternative treatment option in children and adolescents with physical disabilities.

Wang et al. [36] also used six interfaced virtual exercises that are included in the LMC virtual reality system. The games focus on the improvement of dexterity. Their results conclude that LMC facilitates the recovery of the motor function and dexterity of a paretic upper limb. Khademi et al. [37] used LMC to implement the game of Fruit Ninja focusing on finger individuation for stroke patients. The results demonstrated significant correlations between the scores generated from the game and standard clinical outcome measures.

From the review of the literature, it has been observed the importance of the use of contactless systems for certain types of patients who cannot use wearable devices. Likewise, it has been found the effectiveness of therapies that make use of virtual systems as a complement to traditional therapies. Finally, given the different conditions that patients have, it is necessary to adapt the different types of feedback that patients need at the auditory, visual, or tactile level. Even though all of the related works propose different kind of implementations, it had not been found a generic architecture for virtual reality systems oriented to upper limb rehabilitation.

III. METHODS

A. Architecture Proposal Methodology

The software architecture of a system is the structure that considers: 1) software components, 2) the externally visible properties of those components, and 3) the relationships between the components. Software architecture is important because it defines a set of constraints on the subsequent implementation and it focuses on component assembly [38].

For the architecture proposal, this work have adapted the methodology proposed by Parisaca et al. [39] considering only six steps 1) Identification of system quality attribute requirements; 2) Identification of architecturally significant requirements; 3) Design of architecture components; 4) Classification of components; 5) Validation of design decisions; and 6) Analysis and evaluation of software architecture.

The development of each step is described in section B.

B. Multimodal Contactless Architecture for Upper Limb Virtual Rehabilitation

Step 1: Identification of system quality attribute requirements. Upper limbs rehabilitation is important since it allows the use of hands to interact physically in ADL [5]. Sometimes different kinds of medical conditions do not allow the use of wearable devices for virtual rehabilitation. This reason makes it necessary to consider contactless devices for gesture recognition.

Step 2: Identification of architecturally significant requirements. The functional requirements related to the architectural components are shown in Table I.

TABLE I. FUNCTIONAL REQUIREMENTS

Functional Requirements	Architecture Component
Upper Limb interaction	Virtual Reality System: for upper limb rehabilitation
Data capture through hardware	Contactless Tracking Device
Multiple sources of information for feedback to increase participant motivation and improve rehabilitation effectiveness	Visual feedback Auditory feedback Others

Step 3 Design of architecture components.

Based on Dumas, Lalanne and Oviatt [40] work, we propose an architecture for contactless virtual rehabilitation systems that focus on multimodal feedback. The architecture has four components (Fig. 1): i) The Input Modality; ii) The Integration Committee; iii) The Output Modalities; and iv) The Virtual Reality System. The Integration Committee has three elements: i) Dialog Management, ii) Context User Model History, and iii) Output Modalities Fission.

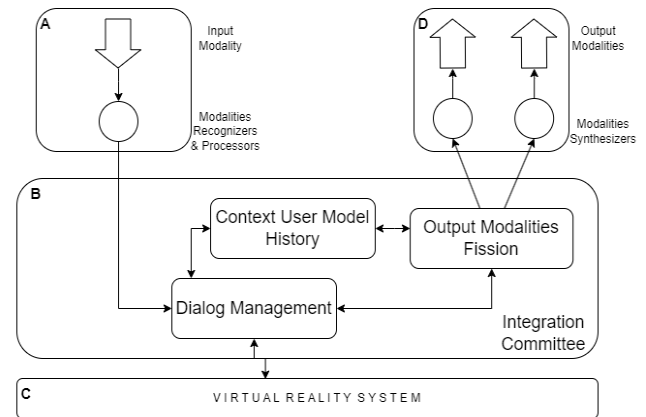


Fig. 1. VRS multimodal contactless architecture.

Input Modality refers to the contactless device that communicates to the Dialog Management. Dialog Management is in charge of identifying the dialog state that takes place in the VRS (the actions to communicate to the VRS and/or the messages to return through the Output Modalities Fission). Output Modalities refers to the auditory or visual feedback that gets the user.

Output Modalities Fission is in charge of returning the feedback to the user through a combination of modalities, depending on the Context User Model History (visual, auditory or other).

Step 4: Classification of components. From the functional requirements described in Table I we have considered: component A (Input Modality) as unimodal and component D (Output Modalities) as multimodal.

Step 5: Validation of the Design Decisions. The validation of the architecture design can be done with different techniques such as scenarios, questionnaires, simulations,

mathematical models, or prototypes [38]. In this work we decided to validate the architecture with a prototype.

For the prototype, the proposed VRS focuses on performing the coordinated actions of handling objects: picking up, manipulating, and releasing them in order to perform exercises to recover hand dexterity [41]. The task of the VRS is to preparing a pizza. The participant must pick up a highlighted ingredient, (one by one) and drop them onto the pizza dough. For each ingredient placed correctly, the participant receives a point (visual feedback). The VRS shows whether it is a hit or a failure (visual feedback). Auditory feedback is also provided, in the case of a hit, a bell rings, and in case of a miss, an error horn sounds (Fig. 2).

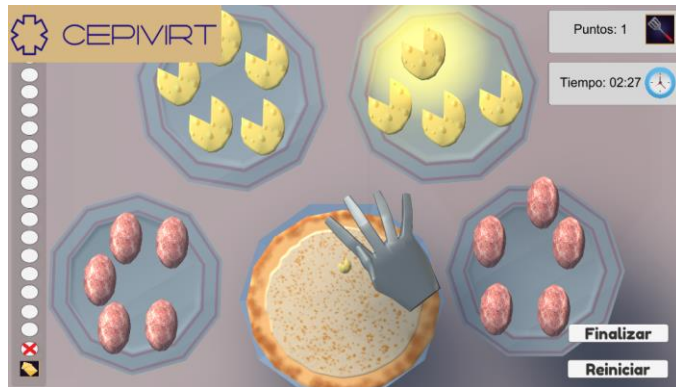


Fig. 2. VRS Pizza Game – right hand – 20 elements.

The VRS allows configuring the hand to be used (left or right) and the number of ingredients to be displayed: 5, 10, 15, or 20.

The VRS was developed in Unity with the contactless optical tracking device Leap Motion Controller (LMC). This is a small optical device with sub-millimeter precision oriented to gestural hand movement [13].

The architecture of the VRS is shown in Fig. 3. The Input Modality considers the contactless device LMC. The Output Modalities are implemented with visual and auditory feedback. We also show the interaction of the VRS states. From the user perspective the Decision state represents the person's attention, intention, and emotions. The Action state represents the hand movements. The Perception state is the recognition of the gestures and movements controlled by the LMC. In the Interpretation state, the data captured from the device is processed.

From the perspective of the VRS, the Computation state performs the fission process that allows the VRS to generate the feedback messages based on the context of the user. The Action state refers to the response to the user action in the form of visual and audio cues. The Perception state refers to what the user hears and sees in the VRS. Finally, the Interpretation state refers to new decisions that the participant will make.

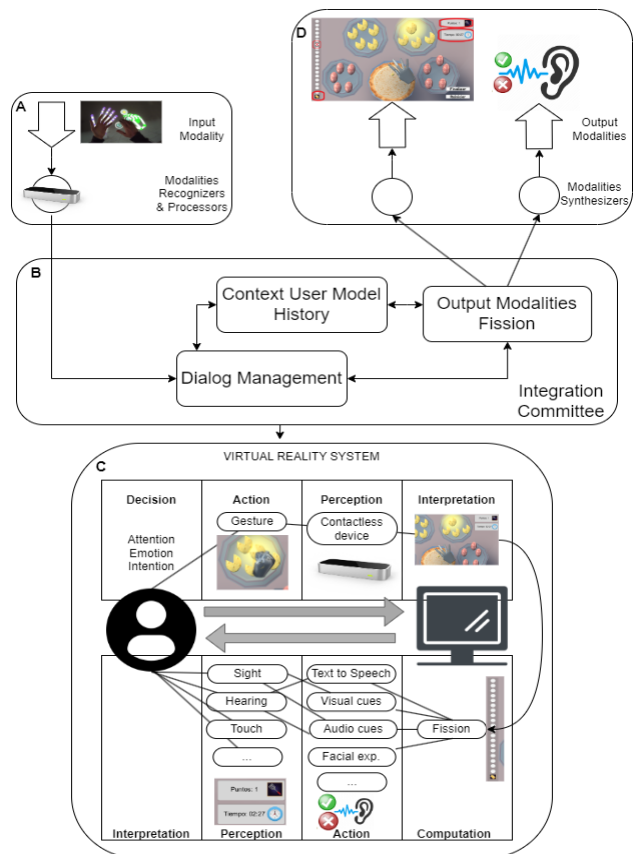


Fig. 3. VRS Pizza Game – Architecture.

Step 6 Analysis and evaluation of software architecture. Two experiments have been performed in order to evaluate the acceptance and technical effectivity of the implementation of the proposed VRS.

Experiment 1

The participants were 106 healthy university students. There were sixty-eight males and thirty-eight females. The mean age of the participants was 20 ± 1.4 years old. The instrument was the User Experience Questionnaire (UEQ) [42,43], which measures six dimensions: (a) Attractiveness: attractive, pleasant, friendly, and enjoyable; (b) Efficiency: to perform tasks quickly, efficiently and pragmatically; (c) Perspicuity: easy to understand, clear, simple, and easy to learn; (d) Reliability: interaction should be predictable, safe and meet user expectations; (e) Stimulation: interesting, exciting, and motivating; (f) Novelty: innovative, inventive, and creatively designed. The scale ranges from -3 to +3. Values between -0.8 and 0.8 represent a more or less neutral evaluation of the corresponding scale, values greater than 0.8 represent a positive evaluation, and values lower than -0.8 represent a negative evaluation [43].

With regards to the protocol, first, the researchers explained the instructions for interacting with the VRS. Then each participant interacted with their dominant hand. Each participant interacted with 20 elements without a time limit. Finally, participants filled out the UEQ questionnaire (Fig. 4 shows the interaction of one student).

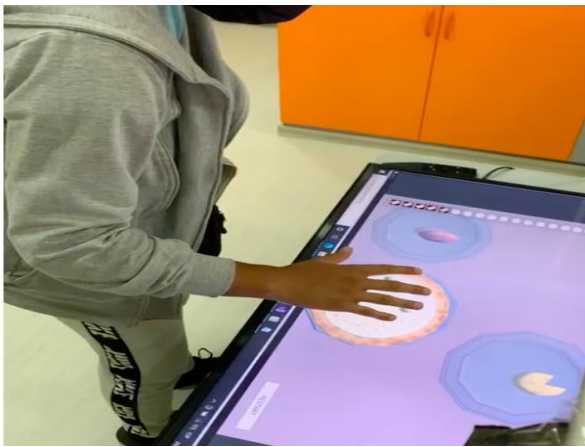


Fig. 4. VRS Pizza Game – Interaction.

Experiment 2

The participants were four children with moderate hand disabilities. There were two males and two females. The mean age of the children was 12.25 ± 3.4 years old. The study followed the guidelines of the Declaration of Helsinki. It was approved by the Ethical Committee (reference number: 2008-0234).

The second intervention was to evaluate the kinematic outcomes of the study. It focused on the number of hits the participants had when they dropped the ingredients onto the dough. With regards to the protocol, the children interacted with the VRS in ten sessions, first with five elements using their right hand and then with their left hand. Then they interacted with the VRS with ten elements using their right hand and then their left hand. The VRS recorded the number of hits in each interaction.

IV. RESULTS AND DISCUSSION

Experiment 1

Table II shows the values of the 6 UEQ scales, all of which have positive results: < 0.8 .

Fig. 5 shows that the highest value is Perspicuity, with an average of 2.068, which is considered to be Excellent. The perception of clarity is based on the comments obtained from the open-ended questions, which highlight the interactivity and ease of use, the intuitiveness of the game, and the feedback channels (visual and audio cues) that are available to the participants.

TABLE II. RESULTS OF THE UEQ SCALES (MEAN AND VARIANCE)

UEQ SCALES (MEAN AND VARIANCE)		
ATTRACTIVENESS	1.987	0.96
PERSPICUITY	2.068	1.31
EFFICIENCY	1.517	1.15
DEPENDABILITY	1.649	0.95
STIMULATION	1.703	1.16
NOVELTY	1.401	1.33

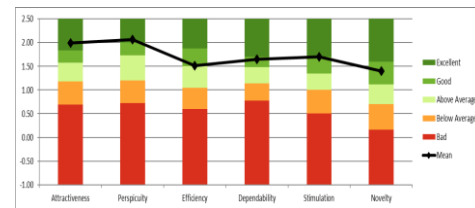


Fig. 5. Results of the six main areas of user experience according to UEQ, in qualitative intervals.

The second highest value was Attraction, with an average of 1.987, which is considered to be Excellent. According to the opinions of the participants, they found the system to be friendly because, by having both visual and audio feedback, they could achieve the objective of the game. A second aspect mentioned was the use of the LMC for hand recognition, which has no physical contact with the user.

The Stimulation obtained a value of 1.703, which is a value between Excellent and Good. The participants found the interaction with the LMC interesting for moving virtual objects. They also highlighted the importance of sounds for hits and faults. In addition, since it is a scoring game, the users want to improve their performance.

Dependability obtained 1.649, which places it on the borderline between Excellent and Good. A requirement to meet this value is the correct functioning of the input and output modalities, which allows the user's expectations to be met in terms of the virtual hand behaving in the same way as the real hand (no delay in execution time). Safety has been guaranteed since the Leap Motion Controller is certified as being compliant with safety and electrical regulatory standards and has no contact with the user.

Efficiency scored a value of 1.517. The system was considered to be fast and efficient because the user's hand movements are reflected in real-time. However, during the tests, the users identified some occasions in which the hand was not visualized and the pizza topping did not end up falling onto the dough.

Finally, Novelty obtained a value of 1.401 which is considered to be as good. Novelty is given by the creative design of the game that refers to an activity of daily life such as preparing a pizza. The novelty is also given by the fact that the game seeks to be applied as a complement to motor skills rehabilitation therapies. The system captures the history of each participant using the time it takes to move each of the ingredients and the number of success/failures. This information allows performance over time to be evaluated.

Experiment 2

Fig. 6 shows the kinematic outcomes. The study analyzed hand dexterity by counting the number of hits per session.

For the interaction using both the right hand and the left hand with five virtual objects (Fig. 6(a) and Fig. 6(b)), there are different performance measures for each participant. We observed better performance in the interaction with the right hand. This is explained by the fact that, for the four participants, the dominant hand was the right hand.

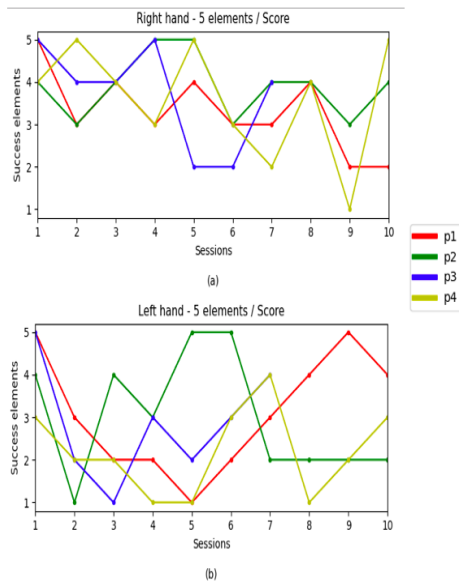


Fig. 6. Results of kinematics: hand dexterity – hits per session.

These results show physical therapist quantitative data in manual dexterity regarding accuracy to capture the dexterity required to complete ADL. This work presents an adapted version of the architecture proposed by Dumas, Lalanne and Oviatt [40] by including only one device for the Input Modality: a contactless device. This decision is proposed based on the context of the type of patients with upper limb disabilities. Likewise, today, contactless devices have shown greater accuracy for data capture. This decision also reduces the complexity described in [40] for the fusion processes in multimodal inputs: data-level fusion and feature-level fusion.

The simplified architecture proposal considered only one input contactless device. This prevents having to deal with problems of noise when dealing with multiple signals. This work considers that upper limb gesture recognitions could be done with only one device since different studies have demonstrated its accuracy [13] [44]. Feature-level fusion handles noise better but needs numerous data training sets before satisfactory performance can be achieved.

V. CONCLUSION

This work has proposed a Multimodal Contactless Architecture for upper limb Virtual Rehabilitation. This work has also implemented a Virtual Reality System based on the architecture for upper limb rehabilitation using the contactless device LMC.

The use of the proposed architecture has allowed the orchestration of four components: 1) the use of a contactless device for gesture recognition; 2) audio and visual cues for multimodal feedback; 3) an integration committee that performs the orchestration between the four components; and 4) the virtual rehabilitation system: Pizza-Game. The architecture focused on multimodal feedback. This study shows that this architecture could be useful for developers for VRS that do not require the use of complex and multiple devices for input modalities. The proposed architecture has considered specific constraints such as the use of contactless

devices for patients that can not use wearable devices due to their medical conditions.

This study had a few limitations. First, the study has only focused on commercial devices that have demonstrated their accuracy such as the Leap Motion Controller device. But the use of custom-made contactless devices whose accuracy has not been proven has not been considered, in that case, an architecture with more input devices is required. Second, since the sample size is small, one must be cautious with respect to the kinematics results obtained at a statistical level. Future studies should be carried out with a representative sample size at a statistical level.

As future work, it is planned to experiment with other contactless devices in order to compare accuracy and user experience. We also plan to incorporate some other devices for multimodal feedback.

ACKNOWLEDGMENT

This contribution was funded by the Universidad Nacional de San Agustín de Arequipa under contract IB-42-2020-UNSA- project "Virtual Rehabilitation System (VR) for motor and cognitive improvement in children with Epileptic Encephalopathy, CEPIVIRT".

REFERENCES

- [1] A. Pollock, S. E. Farmer, M. C. Brady, P. Langhorne, G. E. Mead, J. Mehrholz, and F. van Wijck, "Interventions for improving upper limb function after stroke," *Cochrane Database of Systematic Reviews*, no. 11, 2014.
- [2] I. E. Scheffer and J. Liao, "Deciphering the concepts behind 'Epileptic encephalopathy' and 'Developmental and epileptic encephalopathy,'" *European journal of paediatric neurology*, vol. 24, pp. 11–14, 2020.
- [3] J.-H. Moon, J.-H. Jung, S.-C. Hahm, and H. Cho, "The effects of task-oriented training on hand dexterity and strength in children with spastic hemiplegic cerebral palsy: A preliminary study," *J Phys Ther Sci*, vol. 29, no. 10, pp. 1800–1802, 2017.
- [4] S. Tan, C. T. Hong, J.-H. Chen, L. Chan, W.-C. Chi, C.-F. Yen, H.-F. Liao, T.-H. Liou, and D. Wu, "Hand fine motor skill disability correlates with cognition in patients with moderate-to-advanced Parkinson's disease," *Brain Sci*, vol. 10, no. 6, p. 337, 2020.
- [5] M. Vergara, J. L. Sancho-Bru, V. Gracia-Ibáñez, and A. Pérez-González, "An introductory study of common grasps used by adults during performance of activities of daily living," *Journal of Hand Therapy*, vol. 27, no. 3, pp. 225–234, 2014.
- [6] R. K. Powell and R. L. von der Heyde, "The inclusion of activities of daily living in flexor tendon rehabilitation: a survey," *Journal of Hand Therapy*, vol. 27, no. 1, pp. 23–29, 2014.
- [7] X. Chen, L. Gong, L. Wei, S.-C. Yeh, L. D. Xu, L. Zheng, and Z. Zou, "A wearable hand rehabilitation system with soft gloves," *IEEE Trans Industr Inform*, vol. 17, no. 2, pp. 943–952, 2020.
- [8] D. K. Zondervan, N. Friedman, E. Chang, X. Zhao, R. Augsburger, D. J. Reinkensmeyer, and S. C. Cramer, "Home-based hand rehabilitation after chronic stroke: Randomized, controlled single-blind trial comparing the MusicGlove with a conventional exercise program," *J Rehabil Res Dev*, vol. 53, no. 4, pp. 457–472, 2016.
- [9] Q. Sanders, V. Chan, R. Augsburger, S. C. Cramer, D. J. Reinkensmeyer, and A. H. Do, "Feasibility of wearable sensing for in-home finger rehabilitation early after stroke," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1363–1372, 2020.
- [10] L. Tychsen and L. L. Thio, "Concern of photosensitive seizures evoked by 3D video displays or virtual reality headsets in children: current perspective," *Eye Brain*, pp. 45–48, 2020.

- [11] A. M. Ashleibta, A. Taha, M. A. Khan, W. Taylor, A. Tahir, A. Zoha, Q. H. Abbasi, and M. A. Imran, "5g-enabled contactless multi-user presence and activity detection for independent assisted living," *Sci Rep*, vol. 11, no. 1, p. 17590, 2021.
- [12] Y. Zhu, W. Lu, W. Gan, and W. Hou, "A contactless method to measure real-time finger motion using depth-based pose estimation," *Comput Biol Med*, vol. 131, p. 104282, 2021, doi: <https://doi.org/10.1016/j.combiomed.2021.104282>.
- [13] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.
- [14] J. Guna, G. Jakus, M. Pogačnik, S. Tomažič, and J. Sodnik, "An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking," *Sensors* 2014, Vol. 14, Pages 3702–3720, vol. 14, no. 2, pp. 3702–3720, Feb. 2014, doi: [10.3390/S140203702](https://doi.org/10.3390/S140203702).
- [15] M. N. Eshwarappa and M. V Latte, "Multimodal biometric person authentication using speech, signature and handwriting features," *International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence*, vol. 1, no. 3, pp. 77–86, 2011.
- [16] S. Oviatt, "Multimodal interfaces," *The human-computer interaction handbook*, pp. 439–458, 2007.
- [17] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review," *Psychon Bull Rev*, vol. 20, pp. 21–53, 2013.
- [18] Y. Hernandez-Mella, A. Marin-Hernandez, E. J. Rechy-Ramirez and L. F. Marin-Urias, "A Study of Contactless Human Computer Interaction with Virtual Environments," 2019 5th Experiment International Conference (exp.at'19), Funchal, Portugal, 2019, pp. 16-21, doi: [10.1109/EXPAT.2019.8876576](https://doi.org/10.1109/EXPAT.2019.8876576).
- [19] E. Tarakci, N. Arman, D. Tarakci, and O. Kasapcopur, "Leap Motion Controller-based training for upper extremity rehabilitation in children and adolescents with physical disabilities: A randomized controlled trial," *Journal of Hand Therapy*, vol. 33, no. 2, pp. 220–228, 2020.
- [20] A. Cuesta-Gómez, P. Sánchez-Herrera-Baeza, E. D. Oña-Simbaña, A. Martínez-Medina, C. Ortiz-Comino, C. Balaguer-Bernaldo-de-Quirós, A. Jardón-Huete, and R. Cano-de-la-Cuerda, "Effects of virtual reality associated with serious games for upper limb rehabilitation in patients with multiple sclerosis: Randomized controlled trial," *J Neuroeng Rehabil*, vol. 17, pp. 1–10, 2020.
- [21] E. Avcil, D. Tarakci, N. Arman, and E. Tarakci, "Upper extremity rehabilitation using video games in cerebral palsy: a randomized clinical trial," *Acta Neurol Belg*, vol. 121, pp. 1053–1060, 2021.
- [22] S. I. Afsar, I. Mirzayev, O. U. Yemisci, and S. N. C. Saracgil, "Virtual reality in upper extremity rehabilitation of stroke patients: a randomized controlled trial," *Journal of Stroke and Cerebrovascular Diseases*, vol. 27, no. 12, pp. 3473–3478, 2018.
- [23] C. Francisco-Martínez, J. A. Padilla-Medina, J. Prado-Olivarez, F. J. Pérez-Pinal, A. I. Barranco-Gutiérrez, and J. J. Martínez-Nolasco, "Kinect v2-assisted semi-automated method to assess upper limb motor performance in children," *Sensors*, vol. 22, no. 6, p. 2258, 2022.
- [24] M. I. Daoud, A. Alhusseini, M. Z. Ali, and R. Alazrai, "A game-based rehabilitation system for upper-limb cerebral palsy: a feasibility study," *Sensors*, vol. 20, no. 8, p. 2416, 2020.
- [25] Y. M. Lee, S. Lee, K. E. Uhm, G. Kurillo, J. J. Han, and J. Lee, "Upper limb three-dimensional reachable workspace analysis using the Kinect sensor in hemiplegic stroke patients: A cross-sectional observational study," *Am J Phys Med Rehabil*, vol. 99, no. 5, pp. 397–403, 2020.
- [26] N. Hu, P. H. Chappell and N. R. Harris, "Finger Displacement Sensing: FEM Simulation and Model Prediction of a Three-Layer Electrode Design," in *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 5, pp. 1432-1440, May 2019, doi: [10.1109/TIM.2018.2884545](https://doi.org/10.1109/TIM.2018.2884545).
- [27] A. Jena, J. Chong, A. Jafari and A. Etoundi, "Therapy Easy: A co-designed hand rehabilitation system using Leap motion controller," 2021 24th International Conference on Mechatronics Technology (ICMT), Singapore, 2021, pp. 1-5, doi: [10.1109/ICMT53429.2021.9687286](https://doi.org/10.1109/ICMT53429.2021.9687286).
- [28] M. Alimanova et al., "Gamification of Hand Rehabilitation Process Using Virtual Reality Tools: Using Leap Motion for Hand Rehabilitation," 2017 First IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 2017, pp. 336-339, doi: [10.1109/IRC.2017.76](https://doi.org/10.1109/IRC.2017.76).
- [29] R. Herne, M. F. Shiratuddin, S. Rai, D. Blacker and H. Laga, "Improving Engagement of Stroke Survivors Using Desktop Virtual Reality-Based Serious Games for Upper Limb Rehabilitation: A Multiple Case Study," in *IEEE Access*, vol. 10, pp. 46354-46371, 2022, doi: [10.1109/ACCESS.2022.3169286](https://doi.org/10.1109/ACCESS.2022.3169286).
- [30] J.-D. Huang, "Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities," in *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 2011, pp. 319–320.
- [31] I. Pastor, H. A. Hayes, and S. J. M. Bamberg, "A feasibility study of an upper limb rehabilitation system using kinect and computer games," in *2012 annual international conference of the ieee engineering in medicine and biology society*, 2012, pp. 1286–1289.
- [32] R. J. Adams, M. D. Lichter, E. T. Krepkovich, A. Ellington, M. White, and P. T. Diamond, "Assessing upper extremity motor function in practice of virtual activities of daily living," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 2, pp. 287–296, 2014.
- [33] H. Mousavi Hondori and M. Khademi, "A Review on Technical and Clinical Impact of Microsoft Kinect on Physical Therapy and Rehabilitation," *J Med Eng*, vol. 2014, pp. 1–16, Dec. 2014, doi: [10.1155/2014/846514](https://doi.org/10.1155/2014/846514).
- [34] W. L. Ding, Y. Z. Zheng, Y. P. Su, and X. L. Li, "Kinect-based virtual rehabilitation and evaluation system for upper limb disorders: A case study," *J Back Musculoskelet Rehabil*, vol. 31, no. 4, pp. 611–621, 2018.
- [35] E. Tarakci, N. Arman, D. Tarakci, and O. Kasapcopur, "Leap Motion Controller-based training for upper extremity rehabilitation in children and adolescents with physical disabilities: A randomized controlled trial," *Journal of Hand Therapy*, vol. 33, no. 2, pp. 220–228, 2020.
- [36] Z. Wang, P. Wang, L. Xing, L. Mei, J. Zhao, and T. Zhang, "Leap Motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients," *Neural Regen Res*, vol. 12, no. 11, p. 1823, 2017.
- [37] M. Khademi, H. Mousavi Hondori, A. McKenzie, L. Dodakian, C. V. Lopes, and S. C. Cramer, "Free-hand interaction with leap motion controller for stroke rehabilitation," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, 2014, pp. 1663–1668.
- [38] L. Bass, P. Clements, and R. Kazman, *Software architecture in practice*. Addison-Wesley Professional, 2003.
- [39] E. E. S. Parisaca, S. J. M. Muñoz, E. V. Duarte, E. G. C. Gutierrez, A. Y. C. Peraltilla, and S. A. Peérez, "Dynamic Software Architecture Design for Virtual Rehabilitation System for Manual Motor Dexterity," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 2, pp. 78–85, 2023, doi: [10.14569/IJACSA.2023.0140210](https://doi.org/10.14569/IJACSA.2023.0140210).
- [40] B. Dumas, D. Lalanne, and S. Oviatt, "Multimodal interfaces: A survey of principles, models and frameworks," in *Human machine interaction: Research results of the mmi program*, Springer, 2009, pp. 3–26.
- [41] World Health Organization, "International Classification of Functioning, Disability and Health (ICF)." Aug. 2023. Available: <https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health>.
- [42] "User Experience Questionnaire (UEQ)." May 2023. Available: <https://www.ueq-online.org>
- [43] M. Schrepp, A. Hinderks, and J. Thomaschewski, "Applying the user experience questionnaire (UEQ) in different evaluation scenarios," in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3*, 2014, pp. 383–392.
- [44] P. P. Valentini and E. Pezzuti, "Accuracy in fingertip tracking using Leap Motion Controller for interactive virtual applications," *Int. J. Interact. Des. Manuf.*, vol. 11, no. 3, pp. 641–650, Aug. 2017, doi: [10.1007/s12008-016-0339-y](https://doi.org/10.1007/s12008-016-0339-y).

Attitude Synchronization and Stabilization for Multi-Satellite Formation Flying with Advanced Angular Velocity Observers

Belkacem Kada¹, Khalid Munawar², Muhammad Shafique Shaikh³

Aerospace Engineering Department, King Abdulaziz University, Jeddah, KSA¹

Electrical & Computer Engineering Department, King Abdulaziz University, Jeddah, KSA^{2,3}

Abstract—This paper focuses on two aspects of satellite formation flying (SFF) control: finite-time attitude synchronization and stabilization under undirected time-varying communication topology and synchronization without angular velocity measurements. First, a distributed nonlinear control law ensures rapid convergence and robust disturbance attenuation. To prove stability, a Lyapunov function involving an integrator term is utilized. Specifically, attitude synchronization and stabilization conditions are derived using graph theory, local finite-time convergence for homogeneous systems, and LaSalle's non-smooth invariance principle. Second, the requirements for angular velocity measurements are loosened using a distributed high-order sliding mode estimator. Despite the failure of inter-satellite communication links, the homogeneous sliding mode observer precisely estimates the relative angular velocity and provides smooth control to prevent the actuators of the satellites from chattering. Simulations numerically demonstrate the efficacy of the proposed design scheme.

Keywords—Attitude synchronization; coordinated control; finite-time control; high-order sliding mode observer; inter-satellite communication links; leader-following consensus; switching communication topology

I. INTRODUCTION

Distributed satellite systems (DSS) are an enabling technology for future distributed space missions since they are designed to interact as multi-agent systems for consensus tracking or formation keeping. DSS are deployed at different altitudes and in various configurations (e.g., formations, clusters, swarms, or cancellations) to accomplish distributed space missions like constellation [1], Earth observation [2], remote sensing [3], communication services [4], and meteorology and environmental tasks [5].

Satellite formation flying (SFF) is an attractive concept of DSS flying in prescribed orbits at a fixed separation distance for a given period. This concept enables flexible, reliable, and low-cost space missions [6,7]. However, SFF control requires tight interactions between participating satellites. Relative motion determination is essential for formation keeping and on-orbit reconfiguration. Therefore, SFF systems must meet strict attitude synchronization and tracking requirements before deployment.

The leader-follower-based attitude synchronization has received growing attention in recent years. Several synchronization protocols for SFF systems under fixed and

switching communication topologies have been reported in the literature. Zhou et al. [8] proposed a finite-time control law that guaranteed the coordination of a spacecraft formation under a fixed communication graph. The control law integrated a sliding-mode-based observer, allowing individual satellites to estimate the desired angular velocity. Although the control scheme showed a fast convergence rate, the control torques exhibited chattering effects, which can harm the actuators. Zhao and Jia [9] used a non-singular terminal sliding mode to design a distributed adaptive attitude synchronization algorithm. The chattering of the controllers was avoided using a boundary layer approach. The control algorithm was validated using only a fixed communication graph, which can restrain its ability when switching communication topologies. The attitude synchronization problem for a distributed SFF was investigated by Wang et al. [10]. The authors designed a free-reference attitude control algorithm under communication constraints, demonstrating an asymptotic convergence with an extensive settling time. Zhang et al. [11] discussed the application of integral sliding mode control to the problem of adaptive attitude tracking. The control input was smoothed via the boundary layer method but at the cost of settling time response. Zhang et al. [12] proposed a finite-time attitude synchronization and orbit-tracking control scheme that demonstrated robustness against node failures and torque disturbances. However, the simulations showed a slow convergence rate and high oscillation torques. Combining LQR and robust control approaches, Liu et al. [13] alleviated the effect of nonlinearities and parametric uncertainties on the performance of SFF attitude alignment. The control scheme was validated for a fixed communication topology presenting an asymptotic convergence. Liu et al. [14] designed an adaptive coordinated attitude synchronization control algorithm and a distributed observer for spacecraft formation over a switching network. Although simulations verified the performance of the proposed control scheme, the observer provided an asymptotic convergence to the leader reference attitude, which can degrade formation performance. Lu and Liu [15] further investigated attitude synchronization under switching topologies. The authors proposed a control scheme to guarantee attitude tracking under global and local failures of inter-satellite communication links (ISCLs). However, the consensus protocols showed an asymptotic convergence and required angular speed measurements.

Similarly, Zhang et al. [16] proposed a finite-time distributed attitude synchronization algorithm that requires the availability of the relative angular velocity for all the following satellites, which often leads to a heavy burden of ISCLs and time-consuming control. Zhang et al. [17] used the set theory to develop a new attitude of cooperative control for different SFF structures. Under fixed-time communication topology, the control design worked adequately for the leader-following and leaderless formation structures; however, neither dynamic communication graphs nor external disturbances were considered. Finally, Wei et al. [18] presented a comprehensive overview of the state-of-the-art communication satellite systems in their survey paper. The authors showed that recent publications have focused on designing synchronization techniques for SFF systems with node failures. They concluded that high-accuracy synchronization could be achieved using two-way ISCLs (i.e., undirected communication graphs) and precise observation techniques.

This paper provides solutions to two open problems above finite-time attitude synchronization under ISCLs breaks and robust relative angular velocity estimation. The first problem is solved by designing a distributed attitude synchronization algorithm using a Lyapunov function that involves an integrator term enabling a fast convergence rate. In contrast to the finite-time synchronization schemes available in the literature, the proposed algorithm allows attitude synchronization with time-varying communication graphs. As for the second problem, a high-order sliding mode differentiator estimates the relative angular velocity, guarantees formation robustness, and avoids chattering effects. The distributed observer helps relax the communication topology requirements and reduce the ISCLs burden. Most of the existing velocity-free or observer-based protocols have a slow convergence rate, lack of robustness, or perform only for fixed-time communication topology (e.g., [8,19-23]).

The outline of this paper is given as follows. Section II presents the preliminaries of this work and describes quaternion-based satellite attitude dynamics. A distributed finite-time attitude synchronization controller and relative angular velocity estimator are designed in Section III. Numerical simulations are carried out in Section IV to demonstrate the effectiveness and robustness of the proposed control scheme. Finally, Section V concludes the work.

II. PRELIMINARIES AND DYNAMIC MODELING

A. Graph Theory and Preliminaries

The leader-follower consensus approach is considered in this study to design distributed attitude synchronization and stabilization (ASS) protocols. The satellites are regarded as followers denoted by $'i = 1, 2, \dots, n'$, and the leader (i.e., desired attitude) is represented by $'i = 0'$. The ISCLs topology is modeled by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ where $\mathcal{V} = (v_1, v_2, \dots, v_n)$ denotes the vertices set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the edge set, and $\mathcal{A} = (a_{ij} \geq 0) \in \mathbb{R}^{n \times n}$ is the adjacent matrix corresponding to the graph \mathcal{G} . If a follower-satellite $'i'$ is connected to the leader, then the connection weight is represented by $a_{i0} > 0$; otherwise $a_{i0} = 0$. The

Laplacian matrix $\mathcal{L} = (\ell_{ij}) \in \mathbb{R}^{n \times n}$ associated to the graph \mathcal{G} is defined as

$$\mathcal{L} = \mathcal{D} - \mathcal{A}, \mathcal{D} = \text{diag}(d_{ii}), d_{ii} = \sum_{j \in N_j} a_{ij} \quad (1)$$

where $N_j = \{j | j \in \mathcal{V}, (j, i) \in \mathcal{E}\}$ represents the neighborhood set of a node $'i'$

Lemma 1 [24]: If the matrix \mathcal{L} is associated with a connected undirected graph \mathcal{G} , then its spectrum is given as

- $\lambda_1(\mathcal{L}) = 0$ is a simple eigenvalue of \mathcal{L} with a corresponding eigenvector $\mathbf{1}_n = [1, \dots, 1]_n^T$.
- All nonzero eigenvalues $\lambda_i(\mathcal{L})$ for $1 < i \leq n$ have positive real parts.
- $0 \leq \lambda_2(\mathcal{L}) \leq \dots \leq \lambda_n(\mathcal{L})$.

Lemma 2 [24]: According to the extension theorem, the following condition holds for any integral function f

$$\left\| \int_a^b f(t) dt \right\|_2 \leq (b - a) \|f(t)\|_\infty \quad (2)$$

Definition 1: Let $\mathbf{x} \in \mathbb{R}^n$ be a fixed, measurable function with l_1 -norm $\|\mathbf{x}\|_1 = \max|\mathbf{x}|$ and define an integral operator $L_K \mathbf{x}(t) = \int_X \mathbf{K} \mathbf{x}(t) dt$. Then

$$\|L_K\|_{HS} = \|\mathbf{K}\|_2^2 \quad (3)$$

where X is a compact with kernel \mathbf{K} and HS denotes Hilbert Schmidt norm.

B. Satellite Attitude Kinematics and Dynamics

Let $\bar{\mathbf{q}}_i = [q_{0i} \ \mathbf{q}_i]^T \in \mathbb{R}^4$ and $\boldsymbol{\omega}_i \in \mathbb{R}^3$ denote an i^{th} satellite's orientation and angular velocity within an SFF orbiting in an inertial frame. The nonlinear attitude kinematics and dynamics models are given by

$$\dot{\bar{\mathbf{q}}}_i = -\frac{1}{2} \boldsymbol{\Omega}(\boldsymbol{\omega}) \bar{\mathbf{q}}_i \quad (4)$$

$$\mathbf{J}_i \dot{\boldsymbol{\omega}}_i = -\boldsymbol{\omega}_i \times \mathbf{J}_i \boldsymbol{\omega}_i + \boldsymbol{\tau}_i + \mathbf{d}_i$$

with $\boldsymbol{\Omega}(\boldsymbol{\omega}) \in \mathbb{R}^{4 \times 4}$ is defined as

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} 0 & \omega_3 & -\omega_2 & \omega_1 \\ -\omega_3 & 0 & \omega_1 & \omega_2 \\ \omega_2 & -\omega_1 & 0 & \omega_3 \\ -\omega_1 & -\omega_2 & -\omega_3 & 0 \end{bmatrix}$$

\mathbf{J}_i , $\boldsymbol{\tau}_i$, \mathbf{d}_i denote the inertia tensor, torque vector, and external disturbances vector, respectively. Using the matrix $\mathbf{Q}_i(\mathbf{q}_i) = \mathbf{q}_{0i} \mathbf{I}_{3 \times 3} - \mathbf{q}_i^\times$ the model (4) can be rewritten as follows

$$\begin{aligned} \dot{\mathbf{q}}_i &= -\mathbf{Q}(\mathbf{q}_i) \boldsymbol{\omega}_i \\ \dot{q}_{0i} &= -\frac{1}{2} \boldsymbol{\omega}_i^T \mathbf{q}_i \end{aligned} \quad (5)$$

$$\mathbf{J}_i \dot{\boldsymbol{\omega}}_i = -\boldsymbol{\omega}_i^\times \mathbf{J}_i \boldsymbol{\omega}_i + \boldsymbol{\tau}_i + \mathbf{d}_i$$

where \mathbf{q}_i^\times is the skew-symmetric cross-product matrix and $\mathbf{I}_{3 \times 3}$ denotes a three-dimensional identity matrix. The satellites' angular velocities and external disturbances in the model (5) are bounded, as shown in assumptions 1 and 2 [11].

Assumption 1: $\|\boldsymbol{\omega}_i\|_2 \leq \rho$ and $\|\dot{\boldsymbol{\omega}}_i\|_2 \leq \delta$ with $\rho, \delta \in \mathbb{R}^+$.

Assumption 2: There exist $\gamma_0, \gamma_1 \in \mathbb{R}^+$ such that $\|\mathbf{d}_i\|_2 \leq \gamma_1 \|\boldsymbol{\omega}_i\|_2 + \gamma_0$

The control objective here is to design consensus protocols (control torques $\boldsymbol{\tau}_i$) such that the attitude synchronization of SFF orbiting with time-varying interaction topologies is achieved in a finite time. Thus, $\forall \mathbf{x}_i^0 \in D \subset \mathbb{R}^3$ there exists a synchronization time t_s for which

$$\begin{cases} \lim_{t \rightarrow t_s} \|\mathbf{q}_i - \mathbf{q}_d\|_2 = 0 \\ \lim_{t \rightarrow t_s} \|\boldsymbol{\omega}_i - \boldsymbol{\omega}_d\|_2 = 0 \end{cases} \quad (6)$$

where \mathbf{q}_d and $\boldsymbol{\omega}_d$ denote the desired orientations and angular velocities.

III. SYNCHRONIZATION CONTROL LAW DESIGN

A. SFF Attitude Synchronization under Undirected Time-Varying Interaction Topology

Consider the case of a SFF where at least one follower satellite is connected to the virtual leader. The following finite-time distributed torque law is considered for attitude synchronization of the i^{th} satellite I the SFF

$$\begin{aligned} \boldsymbol{\tau}_i = & \boldsymbol{\omega}_i^\times \mathbf{J}_i \boldsymbol{\omega}_i - \mathbf{J}_i \mathbf{Q}_i \boldsymbol{\omega}_i - \\ & k \mathbf{J}_i \{ [\sum_{j \in \mathcal{N}} a_{ij} (\mathbf{q}_i - \mathbf{q}_j) + a_{i0} (\mathbf{q}_i - \mathbf{q}_d)] + \\ & [\sum_{j \in \mathcal{N}} a_{ij} (\boldsymbol{\omega}_i - \boldsymbol{\omega}_j) + a_{i0} (\boldsymbol{\omega}_i - \boldsymbol{\omega}_d)] \} \end{aligned} \quad (7)$$

with $k \in \mathbb{R}^+$ is a control parameter.

Theorem 1: Consider a satellite formation flying under an undirected time-varying communication graph and suppose that the dynamics described in Eq. (5) are undisturbed and satisfy assumption 1. The control law (7) guarantees the finite-time convergence of the formation satellites' states to the desired attitude and achieves the formation consensus (6) if there exists a positive-definite symmetric matrix \mathbf{M} and a control gain k satisfying the following conditions.

$$\begin{cases} \mathbf{M} = \mathbf{L} + \mathbf{diag}(a_{i0}) \\ k > \frac{1}{\lambda_{\min}(\mathbf{M})} \end{cases} \quad (8)$$

Proof - Define the absolute quaternion and angular velocity tracking errors as follows

$$\begin{cases} \mathbf{q}_{ei} = \mathbf{q}_d^\times \mathbf{q}_i \\ \boldsymbol{\omega}_{ei} = \boldsymbol{\omega}_i - \boldsymbol{\omega}_d \end{cases} \quad (9)$$

Let $\mathbf{q}_e = [\mathbf{q}_{e1}^T, \dots, \mathbf{q}_{en}^T]^T$, $\boldsymbol{\omega}_e = [\boldsymbol{\omega}_{e1}^T, \dots, \boldsymbol{\omega}_{en}^T]^T$, $\mathbf{J} = \mathbf{diag}\{\mathbf{J}_1, \dots, \mathbf{J}_n\}$, and $\mathbf{Q}_e = \mathbf{diag}\{\mathbf{Q}_1, \dots, \mathbf{Q}_n\}$. Using the control law (7), the extension, to n satellites, of the undisturbed form of the torque equation in the model (5) gives

$$\mathbf{J} \dot{\boldsymbol{\omega}}_e = -\mathbf{J} \mathbf{Q}_e \boldsymbol{\omega}_e - k \mathbf{J} (\mathbf{M} \otimes \mathbf{I}_3) (\mathbf{q}_e + \boldsymbol{\omega}_e) \quad (10)$$

where \otimes denotes the Kronecker product and \mathbf{I}_3 denotes the (3x3) identity matrix.

To reach the consensus (6) in finite time, consider the following candidate Lyapunov function with an integrator term

$$V = \frac{1}{2} \left[(\boldsymbol{\omega}_e + \mathbf{q}_e) + \int_0^t (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \right]^T (\mathbf{M} \otimes \mathbf{I}_N) \left[(\boldsymbol{\omega}_e + \mathbf{q}_e) + \int_0^t (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \right] \quad (11)$$

For simplicity, we define a matrix $\bar{\mathbf{M}} = (\mathbf{M} \otimes \mathbf{I}_N)$. Using Eq. (10), the time derivative of the function (11), along with system (5), can be written as

$$\begin{aligned} \dot{V} = & [-k \bar{\mathbf{M}} (\boldsymbol{\omega}_e + \mathbf{q}_e) + (\boldsymbol{\omega}_e + \mathbf{q}_e)]^T \bar{\mathbf{M}} \\ & \left[(\boldsymbol{\omega}_e + \mathbf{q}_e) + \int_0^t (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \right] \\ = & -k (\boldsymbol{\omega}_e + \mathbf{q}_e)^T \bar{\mathbf{M}}^T \bar{\mathbf{M}} (\boldsymbol{\omega}_e + \mathbf{q}_e) \\ & + (\boldsymbol{\omega}_e + \mathbf{q}_e)^T \bar{\mathbf{M}} (\boldsymbol{\omega}_e + \mathbf{q}_e) \\ & - k (\boldsymbol{\omega}_e + \mathbf{q}_e)^T \bar{\mathbf{M}}^T \bar{\mathbf{M}} \int_0^t (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau + \\ & + (\boldsymbol{\omega}_e + \mathbf{q}_e)^T \bar{\mathbf{M}} \int_0^t (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \end{aligned} \quad (12)$$

Since \mathbf{M} is symmetric, then $\bar{\mathbf{M}}^T \bar{\mathbf{M}} = \bar{\mathbf{M}}^2$. With $\lambda_{\min}(\bar{\mathbf{M}}^2) = \lambda_{\min}^2(\bar{\mathbf{M}}) = \lambda_{\min}^2(\mathbf{M})$ and $\lambda_{\max}(\bar{\mathbf{M}}^2) = \lambda_{\max}^2(\mathbf{M}) = \lambda_{\max}^2(\mathbf{M})$, expression (12) can be bounded as follows

$$\begin{aligned} \dot{V} \leq & - (k \lambda_{\min}(\mathbf{M}^2) - \lambda_{\min}(\mathbf{M})) \|\boldsymbol{\omega}_e + \mathbf{q}_e\|_2^2 \\ & - (k \lambda_{\max}(\mathbf{M}^2) - \lambda_{\max}(\mathbf{M})) \|(\boldsymbol{\omega}_e + \mathbf{q}_e)^T\|_1 \|(\boldsymbol{\omega}_e + \mathbf{q}_e)^T\|_\infty \\ \leq & -\lambda_{\min}(\mathbf{M}) (k \lambda_{\min}(\mathbf{M}) - 1) \|\boldsymbol{\omega}_e + \mathbf{q}_e\|_2^2 \\ & - \lambda_{\max}(\mathbf{M}) (k \lambda_{\max}(\mathbf{M}) - 1) \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_1 \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_\infty \end{aligned} \quad (13)$$

Thus if $k > 1/\lambda_{\min}(\mathbf{M})$, then

$$\dot{V} \leq -\lambda_{\min}(\mathbf{M}) (k \lambda_{\min}(\mathbf{M}) - 1) \|\boldsymbol{\omega}_e + \mathbf{q}_e\|_2^2 \quad (14)$$

It follows that $\dot{V} \leq 0$. To prove that V will decrease to zero in finite time, \dot{V} will also satisfy

$$\dot{V} \leq -\lambda_{\max}(\mathbf{M}) (k \lambda_{\max}(\mathbf{M}) - 1) \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_1 \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_\infty \quad (15)$$

According to lemma 2, with the use of equation (11), the function V satisfies

$$\begin{aligned} V \leq & \frac{\lambda_{\max}(\mathbf{M})}{2} \left\| (\boldsymbol{\omega}_e + \mathbf{q}_e) + \int_0^{t_s} (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \right\|_2^2 \\ \leq & \frac{\lambda_{\max}(\mathbf{M})}{2} \left(\|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_2^2 + \left\| \int_0^{t_s} (\boldsymbol{\omega}_e + \mathbf{q}_e) d\tau \right\|_2^2 \right) \\ \leq & \frac{\lambda_{\max}(\mathbf{M})}{2} \left(\|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_1^2 + t_s^2 \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_\infty^2 \right) \\ \leq & \frac{t_s^2 \lambda_{\max}(\mathbf{M})}{2} \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_1^2 \|(\boldsymbol{\omega}_e + \mathbf{q}_e)\|_\infty^2 \end{aligned} \quad (16)$$

where t_s denotes the settling time. It results from Eq. (15) and (16) that

$$\dot{V} \leq -\lambda_{\max}(\mathbf{M}) \left(\frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} - 1 \right) \frac{\sqrt{2}\sqrt{V}}{\sqrt{t_s \lambda_{\max}(\mathbf{M})}} = -\alpha \sqrt{V} \quad (17)$$

which implies that for $\alpha = \frac{\sqrt{2}}{\sqrt{t_s}} \sqrt{\lambda_{\max}(\mathbf{M})} \left(\frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} - 1 \right) > 0$, $V(t) \rightarrow 0$ as $t \rightarrow t_s$.

Finally, the settling time t_s is computed from the integration of Eq. (17) as follows

$$\sqrt{V(t)} \leq \sqrt{V(0)} - \frac{\alpha t_s}{2} \quad (18)$$

Using equation (11) at $t = 0$, the settling time t_s is given as

$$t_s = \frac{\sqrt{2(\boldsymbol{\omega}_e(0) + \mathbf{q}_e(0))^T (\mathbf{M} \otimes \mathbf{I}_N) (\boldsymbol{\omega}_e(0) + \mathbf{q}_e(0))}}{\sqrt{\lambda_{\max}(\mathbf{M})} \left(\frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} - 1 \right)} \quad (19)$$

Thus, $\mathbf{q}_i \rightarrow \mathbf{q}_d$ and $\boldsymbol{\omega}_i \rightarrow \boldsymbol{\omega}_d$ in finite time for $t \geq t_s$.

End of proof.

B. Attitude Synchronization with High-Order Sliding Mode Estimator

In this subsection, a distributed angular velocity observer is introduced to reduce the number of ISCLs and guarantee the robustness of the SFF against loss of communication. First, a finite-time high-order sliding mode differentiator is designed to provide follower satellites with an accurate estimate of the desired angular velocity $\hat{\boldsymbol{\omega}}_i$. To do so, we define the following non-empty sliding mode function.

$$\mathbf{s}(t) = (\mathbf{s}_1^T, \mathbf{s}_2^T, \dots, \mathbf{s}_n^T)^T \quad (20)$$

where $\mathbf{s}_i \in \mathbb{R}^3$ denotes a sliding mode manifold for satellite 'i'. The sliding surface \mathbf{s}_i is defined by an integral sliding mode function as

$$\mathbf{s}_i = \mathbf{e}_{\omega,i} + \int_0^t \mathbf{e}_{\omega,i}(\tau) d\tau \quad (21)$$

with

$$\mathbf{e}_{\omega,i} = \sum_{j \in \mathcal{N}} [a_{ij}(\boldsymbol{\omega}_i - \boldsymbol{\omega}_j) + a_{i0}(\boldsymbol{\omega}_i - \boldsymbol{\omega}_d)] \quad (22)$$

Lemma 1 [25]: Consider a continuous function $f(t)$ with the time-derivative $f^{(r)}(t)$ has a Lipschitz constant $L > 0$ (i.e., $f^{(r)}(t) \leq L$), where r is the relative degree of $\mathbf{s}(t)$. The following high-order sliding mode differentiator produces accurate estimations of $f(t)$ and its successive time derivatives $f^{(k)}(t)$ ($k = 1, \dots, r - 1$) in finite time

$$\begin{cases} \dot{z}_0 = z_1 + \mu_0 |z_0 - f(t)|^{\frac{r-1}{r}} \\ \dot{z}_1 = z_2 + \mu_1 |z_0 - f(t)|^{\frac{r-2}{r-1}} \\ \vdots \\ \dot{z}_{r-2} = z_{r-1} + \mu_{r-2} |z_0 - f(t)|^{\frac{1}{2}} \\ \dot{z}_{r-1} = z_r + \mu_{r-1} \text{sign}(z_0 - f(t)) \end{cases} \quad (23)$$

where z_i ($i = 0, \dots, r$) denote the differentiator states and μ_i define the differentiator parameters. With $z_r = 0$, the differentiator (23) guarantees that $z_0 \rightarrow \hat{f}_i$ and $z_k \rightarrow \hat{f}_i^{(k)}$ ($k = 1, \dots, r - 1$) converge in finite time.

With $f(t) = s_{i,j}(t)$ where 'i' = 1, ..., n' denotes satellite 'i' and 'j' = 1, ..., 3' denotes motion direction, $\hat{\boldsymbol{\omega}}_{i,j} = z_0$ and $\hat{\boldsymbol{\omega}}_{i,j}^{(k)} = z_i^{(k)}$ ($k = 1, \dots, r - 1$) converge in finite time. We note that $\hat{\boldsymbol{\omega}}_{i,j}^{(k)}$ are the successive time-derivatives of the estimate $\hat{\boldsymbol{\omega}}_{i,j}^{(k)}$.

Second, under the results above, the control law (7) is redesigned as follows

$$\begin{aligned} \boldsymbol{\tau}_i = & \boldsymbol{\omega}_i^\times \mathbf{J}_i \boldsymbol{\omega}_i - \mathbf{J}_i \mathbf{Q}_i [\hat{\boldsymbol{\omega}}_i + \sum_{k=1}^{r-2} \mathbf{Q}_i^{(k)} \hat{\boldsymbol{\omega}}_i^{(k)} + \\ & k \{ [\sum_{j \in \mathcal{N}} a_{ij} (\mathbf{q}_i - \mathbf{q}_j) + a_{i0} (\mathbf{q}_i - \mathbf{q}_d)] \\ & + [\sum_{j \in \mathcal{N}} a_{ij} (\hat{\boldsymbol{\omega}}_i - \hat{\boldsymbol{\omega}}_j) + a_{i0} (\hat{\boldsymbol{\omega}}_i - \boldsymbol{\omega}_d)] \}] \end{aligned} \quad (24)$$

where $\mathbf{Q}_i^{(k)}$ denotes the k^{th} time-derivative of the matrix \mathbf{Q}_i used in (5).

Theorem 2: Consider system (5) with $\mathbf{d}_i = 0$ and a sliding function of the form (21). If there exists a constant $k > 0$, then the protocols (24) guarantee that $\lim_{t \rightarrow T} (\mathbf{q}_i - \mathbf{q}_j) = 0$ and $\lim_{t \rightarrow T} (\boldsymbol{\omega}_i - \boldsymbol{\omega}_j) = \lim_{t \rightarrow T} (\boldsymbol{\omega}_i - \boldsymbol{\omega}_d) = 0$, where T is the synchronization time.

Proof: Let $e_0 = z_0 - f(t)$ be the differentiation error, then system (23) can be written as

$$\begin{cases} \dot{e}_0 = e_1 - \mu_0 |e_0|^{-\frac{1}{r}} e_0 \\ \dot{e}_1 = e_2 - \mu_1 |e_0|^{-\frac{1}{r-1}} e_0 \\ \vdots \\ \dot{e}_{r-2} = e_{r-1} - \mu_{r-2} |e_0|^{-\frac{1}{2}} e_0 \\ \dot{e}_{r-1} = -\mu_{r-1} |e_0|^{-1} e_0 - f^{(r)} \end{cases} \quad (25)$$

The error dynamics (25) can be set in the following pseudo linear system form

$$\dot{\mathbf{e}} = \mathbf{A} \mathbf{e} + \mathbf{b} f^{(r)} \quad (26)$$

with

$$\mathbf{e} = [e_0 \ e_1 \ \dots \ e_{r-1}]^T, \quad \mathbf{b} = [0 \ 0 \ \dots \ 1]^T$$

$$\mathbf{A} = \begin{bmatrix} -\mu_0 |e_0|^{-\frac{1}{r}} & 1 & 0 & \dots & 0 \\ -\mu_1 |e_0|^{-\frac{1}{r-1}} & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mu_{r-2} |e_0|^{-\frac{1}{2}} & 0 & 0 & \dots & 1 \\ -\mu_{r-1} |e_0|^{-1} & 0 & 0 & \dots & 0 \end{bmatrix}$$

One can easily find that the determinant of matrix \mathbf{A} is given as

$$\det(\mathbf{A}) = (-1)^r \mu_{r-1} |e_0|^{-1} = \prod_{i=1}^r \bar{\lambda}_i \quad (27)$$

and its characteristic polynomial as

$$p_A(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = \sum_{k=1}^r \lambda^{r-k} (-1)^k \text{tr}(\wedge^k \mathbf{A}) \quad (28)$$

where λ denotes an eigenvalue of \mathbf{A} and $(\wedge^k \mathbf{A})$ denotes the trace of the k^{th} exterior power of \mathbf{A} . Equation (28) gives

$$p_A(\lambda) = \lambda^r + \mu_0 |e_0|^{-\frac{1}{r}} \lambda^{r-1} + \mu_1 |e_0|^{-\frac{1}{r-1}} \lambda^{r-2} + \dots + \mu_{r-1} |e_0|^{-1} \quad (29)$$

The polynomial (29) is stable if its coefficients satisfy the following Schur condition

$$1 > \mu_0 |e_0|^{-\frac{1}{r}} > \mu_1 |e_0|^{-\frac{1}{r-1}} > \dots > \mu_{r-1} |e_0|^{-1} > 0 \quad (30)$$

To fulfill condition (30), the differentiator coefficients μ_i are computed using the following tuning recursive scheme

$$\begin{cases} \mu_0 := \alpha L^{1/r} \\ \mu_i := \mu_{i-1} L^{1/(r-i)} \quad (i = 1, \dots, r - 1) \end{cases} \quad (31)$$

where $\alpha \in \mathbb{R}^+$ is a tuning parameter. According to the scheme (31), for any $r > 1$ and Lipschitz constant $L \in \mathbb{R}^+$, $\exists \alpha$ such that $\mu_0 < 1$. Consequently, condition (30) is guaranteed, and matrix A is stable.

End of proof.

IV. SIMULATION

In this section, numerical simulations are performed to confirm the above theoretical results and prove the performance and effectiveness of the proposed control scheme. The distributed torque laws (7) and (24) and the observer (23) were applied to a four-satellite formation that runs under an undirected switching communication topology shows in Fig. 1. Table I gives the moments of inertia and the initial conditions of the four satellites.

TABLE I. PARAMETERS OF THE SATELLITES

Sat.	Initial attitude/angular velocity (rad/s)	Moment of inertia (kg.m ²)
1	$\mathbf{q}_1 = [0.5916, 0.6, -0.5, -0.2]$ $\boldsymbol{\omega}_1 = [0.02, -0.01, -0.02]$	$\mathbf{J}_1 = \text{diag}[10.15, 10.20, 9.85]$
2	$\mathbf{q}_2 = [0.7874, 0.5, -0.3, -0.2]$ $\boldsymbol{\omega}_2 = [0.03, -0.01, -0.01]$	$\mathbf{J}_2 = \text{diag}[12.1, 10.90, 10.50]$
3	$\mathbf{q}_3 = [0.6245, 0.3, -0.6, -0.4]$ $\boldsymbol{\omega}_3 = [0.02, 0.00, -0.03]$	$\mathbf{J}_3 = \text{diag}[9.55, 12.30, 10.20]$
4	$\mathbf{q}_4 = [0.6403, 0.5, -0.3, -0.5]$ $\boldsymbol{\omega}_4 = [-0.01, 0.02, -0.03]$	$\mathbf{J}_4 = \text{diag}[10.50, 11.20, 10.20]$

First, the satellites are required to align their attitudes to the desired $\bar{\mathbf{q}}_d = (0, 0, 0, 1)^T$ and $\boldsymbol{\omega}_d = (0, 0, 0)^T$. The distributed torque law (7) is applied with $k = 3.7$, and the simulation is run under the communication topology shown in Fig. 1 for a 60 s with a dwell time $\tau = 15$ s. Fig. 2 and 3 depict the quaternions and angular velocities tracking errors, and Fig. 4 shows the required control torques.

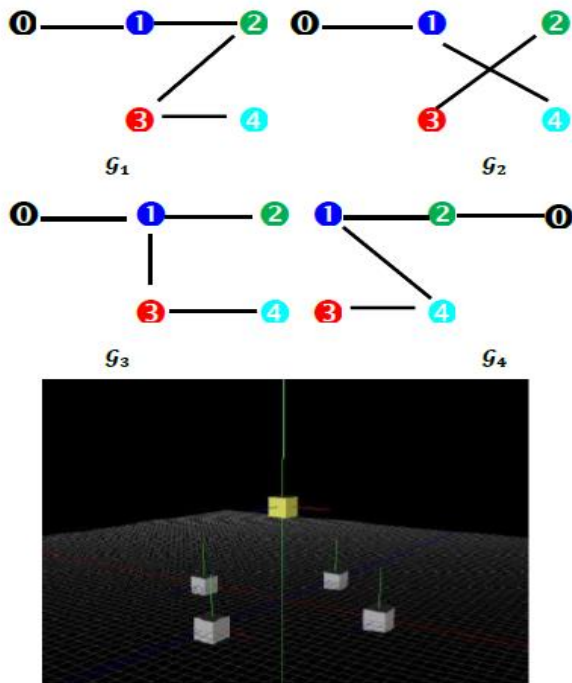


Fig. 1. Switching interaction topology among a four-satellite formation.

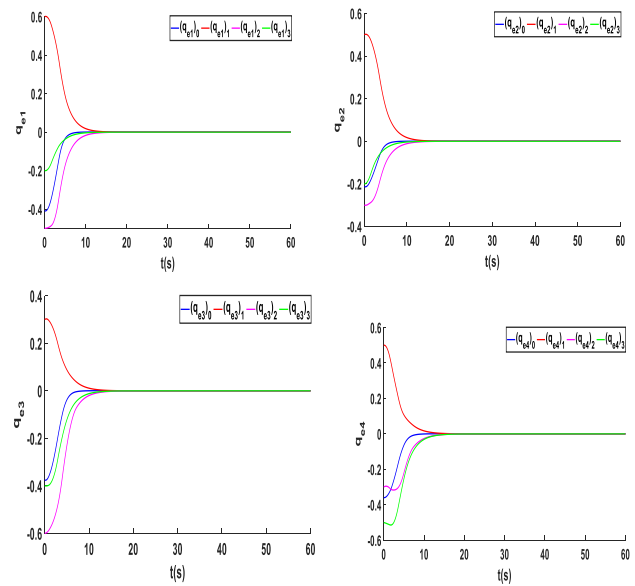


Fig. 2. Quaternions tracking errors.

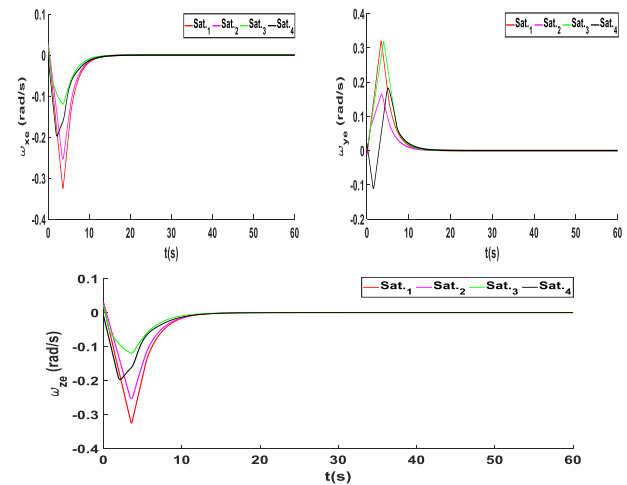


Fig. 3. Angular velocity tracking errors.

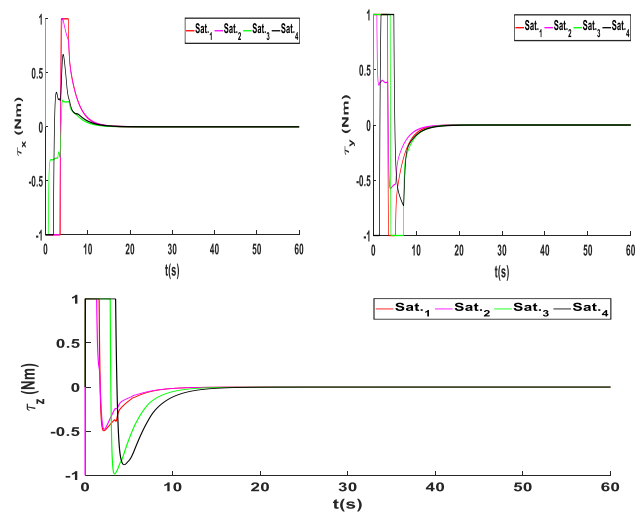


Fig. 4. Control efforts (torques).

The results above indicate that the four satellites can align their attitude to the specified one while effectively maintaining the consensus (6) despite the ISCLS' failures. Next, the case of attitude synchronization without angular velocity measurement is considered. In this scenario, the efficiency of the redesigned torque law (24) with a third-order observer (23) is verified. An external disturbance of $\mathbf{d} = 10^{-4}[\sin(0.12t), \sin(0.5t), \sin(0.18t)]^T$ is introduced to the system (5). The simulation is run with $k = 1.2, \alpha = 0.15$, under the leaderless communication topology shown in Fig. 5. Fig. 6 shows the relative attitude $\mathbf{q}_{ij} = \mathbf{q}_j^* \mathbf{q}_i$ where \mathbf{q}_j^* denotes the cross-product matrix associated with the quaternion vector \mathbf{q}_j . The relative angular velocity errors $\boldsymbol{\omega}_{ei} = \boldsymbol{\omega}_i - \boldsymbol{\omega}_j$ are depicted in Fig. 7.

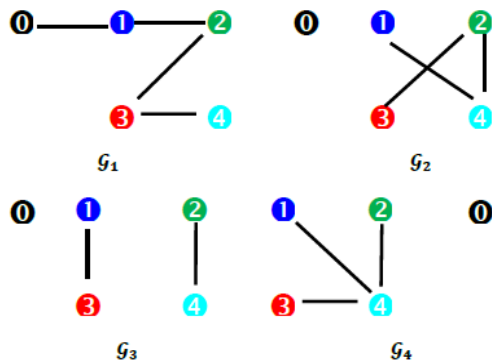


Fig. 5. Switching interaction topology among a four-satellite formation with loss of the leader.

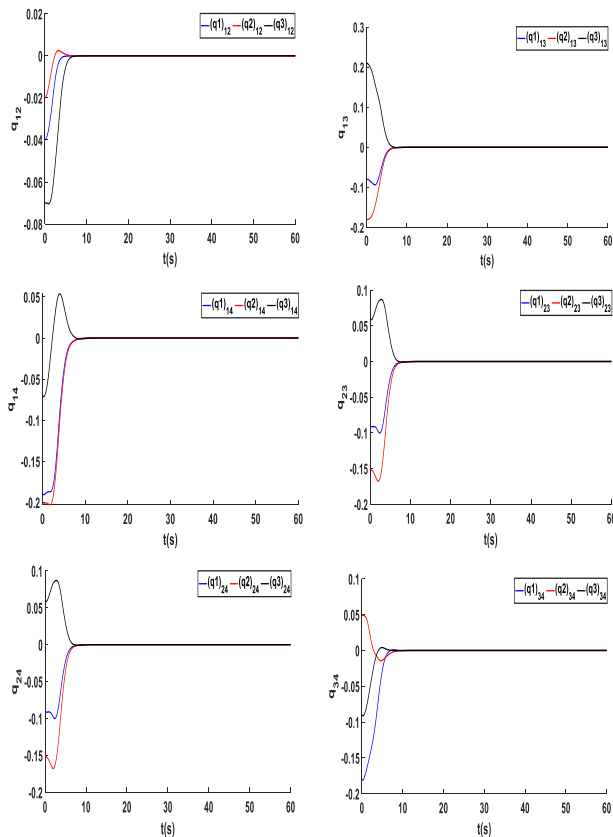


Fig. 6. Relative satellite-to-satellite attitude errors.

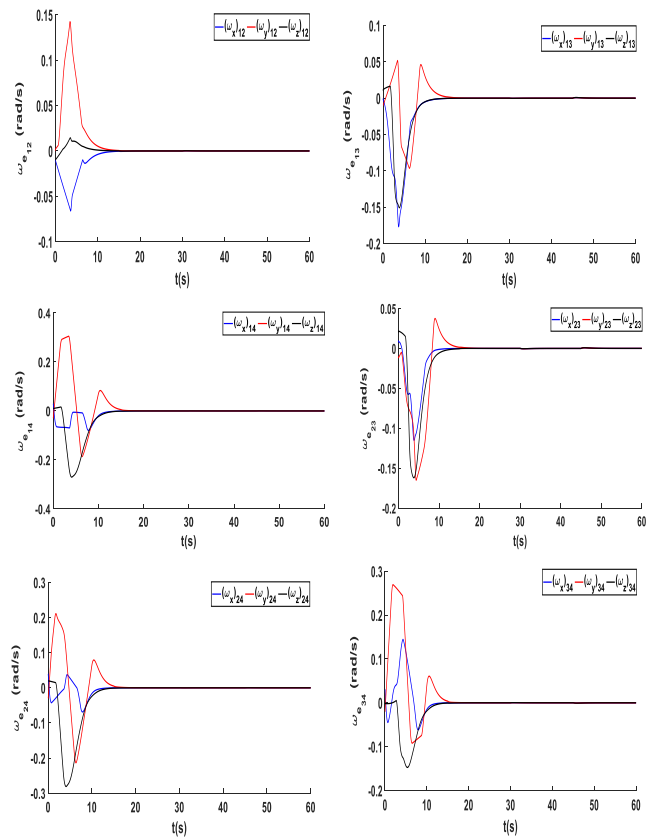


Fig. 7. Relative satellite-to-satellite angular speed errors.

For a fair comparison, attitude synchronization paths for satellites using second and third-order sliding mode observers are shown in Fig. 8.

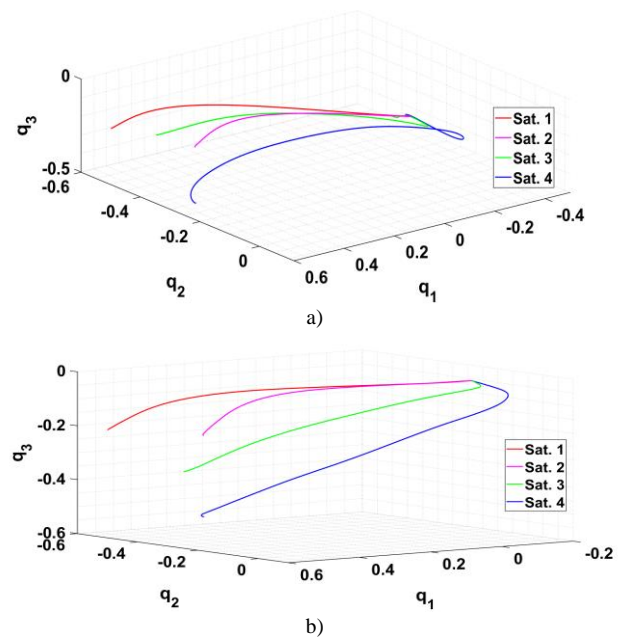


Fig. 8. Attitude synchronization paths with a) second-order observer, b) third-order observer.

Using both static and dynamic communication topologies with and without velocity measurements, the results of the present research demonstrated the feasibility and effectiveness of the proposed finite-time attitude synchronization for satellite formation systems. The results showed that instead of using the actual measured angular velocities, the designed observer with a finite-time high-order sliding-mode generated high-precision estimates. The observer was implemented to lessen the load on the inter-satellite communication lines and guarantee the formation's resilience in the face of link failures. In addition, the performance of the augmented control law (24) is shown to be satisfactory by comparing Fig. 2 and 3 and Fig. 6 and 7. This is the case despite the fact that estimations of relative angular velocities are utilized, as well as the fact that external disturbances are implemented. In addition, it can be shown from Fig. 8 that raising the order of the observer can result in an increase in the precision of the attitude synchronization.

V. CONCLUSION

This paper has investigated the finite-time attitude synchronization for satellite formation systems under switching communication topologies with and without velocity measurements. First, Lyapunov-based distributed torque laws were designed to provide a coordinated synchronization of the states of following satellites with the desired attitudes and angular speeds. The distributed protocols were developed using graph theory, local finite-time convergence for homogeneous systems, and non-smooth LaSalle's invariance principle. Then, a finite-time high-order sliding-mode observer was designed to provide relative angular velocity estimates instead of the measured ones. The distributed observer is introduced to reduce the inter-satellite communication links burden and ensure formation robustness against their breaks. Future works will focus on solving some critical inter-satellite transmission problems such as signal overriding, signal interferences, and communication delays.

ACKNOWLEDGMENT

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. (G:305-135-1441). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

REFERENCES

- [1] S. Bandyopadhyay, R. Foust, G.P. Subramanian, S-J Chung, F.Y Hadaegh, Review of formation flying and constellation missions using nanosatellites, *Journal of Spacecraft and Rockets* 53 (3) (2016) 567–578. <https://doi.org/10.2514/1.A33291>
- [2] J. Sullivan, S. Grimberg, S. D'Amico, Comprehensive survey and assessment of spacecraft relative motion dynamics models, *Journal of Guidance, Control, and Dynamics*. 40 (8) (2017) 1-23. <https://doi.org/10.2514/1.G002309V>
- [3] VLS. Freitas, F.L. de Sousa, E.E.N. Macau, Reactive model for autonomous vehicles formation following a mobile reference, *Applied Mathematical Modelling*. 61 (2018) 167-180. <https://doi.org/10.1016/j.apm.2018.04.011>
- [4] L.M. Marrero, et al., Architectures and synchronization techniques for distributed satellite systems: a survey, *IEEE Access*. (2022). <https://doi.org/10.48550/arXiv.2203.08698>
- [5] O. Mechali, L. Xu, X. Xie, J. Iqbal, Theory and practice for autonomous formation flight of quadrotors via distributed robust sliding mode control

- protocol with fixed-time stability guarantee, *Control Engineering Practice*. 123 (2022) 105150 <https://doi.org/10.1016/j.conengprac.2022.105150>
- [6] D. Wang, B. Wu, and E.K., Poh, Satellite formation flying: Relative dynamics, formation design, fuel optimal maneuvers and formation maintenance, Springer (2017), DOI 10.1007/978-981-10-2383-5
- [7] S. Mathavaraj and R.Padhi, Satellite formation flying: High precision guidance using optimal and adaptive control techniques, Springer (2021) <https://doi.org/10.1007/978-981-15-9631-5>
- [8] J. Zhou, et al., Decentralized finite time attitude synchronization control of satellite formation flying, *Journal of Guidance, Control, and Dynamics* 36 (1) (2013) 185–195. <https://doi.org/10.2514/1.56740>
- [9] L. Zhao, Y. Jia, Decentralized adaptive attitude synchronization control for spacecraft formation using nonsingular fast terminal sliding mode, *Nonlinear Dynamics* 78 (2014) 2779–2794. DOI:10.1007/s11071-014-1625-5
- [10] Z. Wang, et al., Decentralized attitude synchronization of spacecraft formations under complex communication topology, *Advances in Mechanical Engineering* 8 (8) (2016) 1–12. <https://doi.org/10.1177/1687814016660464>
- [11] J. Zhang, Q. Hu, W. Xie, Integral sliding mode-based attitude coordinated tracking for spacecraft formation with communication delays, *International Journal of Systems Science* 48 (15) (2017) 3254-3266. <https://doi.org/10.1080/00207721.2017.1371359>
- [12] J. Zhang, D. Ye, J.D. Biggs, Z. Sun, Finite-time relative orbit-attitude tracking control for multi-spacecraft with collision avoidance and changing network topologies, *Advances in Space Research* 63 (2019) 1161-1175. <https://doi.org/10.1016/j.asr.2018.10.037>
- [13] H. Liu, Y. et al., Robust formation flying control for a team of satellites subject to nonlinearities and uncertainties, *Aerospace Science and Technology*. 95 (2019) 105455. <https://doi.org/10.1016/j.ast.2019.105455>
- [14] X. Liu., Y. Zou, Z. You, Coordinated attitude synchronization and tracking control of multiple spacecraft over a communication network with a switching topology, *IEEE Transactions on Aerospace and Electronic Systems* 56 (2) (2020) 1148-1162. DOI: 10.1109/TAES.2019.2925512
- [15] M. Lu L. Liu, Leader-following attitude consensus of multiple rigid spacecraft systems under switching networks, *IEEE transactions on automatic control* 65 (2) (2020) 839-845. DOI: 10.1109/TAC.2019.2920074
- [16] X. Zhang, et al., Finite-time distributed attitude synchronization for multiple spacecraft with angular velocity and input constraints, *IEEE Transactions on Control Systems Technology* (2021). DOI: 10.1109/TCST.2021.3115479
- [17] S. Zhang, et al., General attitude cooperative control of satellite formation by set stabilization, *Acta Astronautica* 191 (2022) 125-133. <https://doi.org/10.1016/j.actaastro.2021.10.043>
- [18] C. Wei, et al., An overview of prescribed performance control and its application to spacecraft attitude system, *Journal of Systems and Control Engineering* (2020) 1-3. <https://doi.org/10.1177/0959651820952552>
- [19] Q. Hu, J. Zhang, Y. Zhang, Velocity-free attitude coordinated tracking control for spacecraft formation flying, *ISA Transactions* 73 (2018) 54-65, <https://doi.org/10.1016/j.isatra.2017.12.019>
- [20] M. Xu, Y. et al., Chattering free distributed consensus control for attitude tracking of spacecraft formation system with unmeasurable angular velocity, *International Journal of Control, Automation, and Systems*, 18 (9), (2020), 2277-2288. <http://dx.doi.org/10.1007/s12555-019-0543-1>
- [21] H. Yi, Y. Jia, Adaptive finite time distributed 6-DOF synchronization control for spacecraft formation without velocity measurement, *Nonlinear Dynamics* 95 (2019), 2275–2291. <https://doi.org/10.1007/s11071-018-4691-2>
- [22] M. Liu, A. Zhang, B. Xiao, Velocity-free state feedback fault-tolerant control for satellite with actuator and sensor faults, *Symmetry* 14 (157). <https://doi.org/10.3390/sym14010157>
- [23] H. Yang, X. You, C. Hua, Attitude tracking control for spacecraft formation with time-varying delays and switching topology, *Acta*

- Astronautica 126 (2016) 98-108. <https://doi.org/10.1016/j.actaastro.2016.04.012>
- [24] R. Horn, C. Johnson, *Topics in Matrix Analysis*, Cambridge University Press (1991), Cambridge.
- [25] A. Levant, Higher order sliding modes differentiation and output-feedback control, *International Journal of Control*. 76 (9/10) (2003) 924–941. <https://doi.org/10.1080/0020717031000099029>

Hussein Search Algorithm: A Novel Efficient Searching Algorithm in Constant Time Complexity

Omer H Abu El Haijia¹, Arwa H. F. Zabian²
University of Castilla–La Mancha, Toledo-España¹
Jadara University, Jordan-Irbid²

Abstract—Hussein search algorithm focuses on the fundamental concept of searching in computer science and aims to enhance the retrieval of data from various data warehouses. The efficiency of cloud systems is substantially influenced by the manner in which data is saved and retrieved, given the vast quantity of data being generated and stored in the cloud. The act of searching entails the systematic endeavor of locating a particular item within a substantial volume of data, and searching algorithms offer methodical strategies for accomplishing this task. There exists a wide array of searching algorithms, each exhibiting variations in terms of the search procedure, time complexity, and space complexity. The choice of the suitable algorithm is contingent upon various aspects, including the magnitude of the dataset, the distribution of the data, and the desired temporal and spatial intricacy. This study presents a novel prediction-based searching algorithm named the Hussein search algorithm. The system is designed to operate in a straightforward manner and makes use of a simple data structure. This study relies on fundamental mathematical computations and incorporates the interpolation search algorithm, an algorithm that introduces a search by-prediction method for uniformly distributed lists, it forecasts the precise position of the queried object. The cost of prediction remains consistent and, in numerous instances, falls under the $O(1)$ range. Hussein search algorithm exhibits enhanced efficiency in comparison to the binary search and ternary search algorithms, both of which are widely regarded as the best methods for searching sorted data.

Keywords—Binary search; prediction search procedure; prediction cost; constant time complexity

I. INTRODUCTION

On a daily basis, a substantial volume of data is generated across many formats, including photographs, videos, and text. This material is subsequently stored in cloud-based repositories, serving as a collective resource for retrieval from any given database. The increasing complexity of this matter can be attributed to the substantial volume of data that is generated and stored on a daily basis. The act of searching is of significant importance in various contexts, regardless of whether the item being sought is stored within a cloud-based infrastructure or a localized database. In both scenarios, the use of a search algorithm is essential for the successful retrieval of the desired item. The process of searching is commonly employed as a means of problem-solving, whereby the problem is provided as input and a solution is generated in the form of a sequential set of activities. Numerous instances in practical contexts can be classified as searching problems, such as the task of determining the shortest route between two

nodes. These types of problems can be effectively addressed through the use of graph search algorithms. The act of searching can be categorized into two main types: sequential search and binary search. Various searching algorithms exist, each employing distinct strategies and exhibiting variations in terms of time and space complexity. Certain algorithms employ an informed approach, while others adopt a uniform approach, and a third category utilizes a partial information strategy for the purpose of item retrieval. The uniformity of the sequential search method arises from its lack of concern for any prior knowledge regarding the distribution of items. The binary search algorithm is considered an informed search algorithm due to its reliance on a sorted array. In recent years, there has been limited progress in enhancing the complexity of the search algorithm. This is primarily due to the satisfactory performance of the binary search algorithm in terms of searching complexity. However, it is important to note that the binary search algorithm does encounter challenges related to sorting, as it can only operate on sorted arrays. Additionally, it faces difficulties when dealing with comparison-based input involving searching for candidate items. Hence, the temporal complexity is intricately linked to the duration required for the sorting process, resulting in a trade-off where the time saved when searching is offset by the time invested in sorting. The interpolation search algorithm has a temporal complexity of $O(1)$ when the items in the list are evenly distributed [1]. Therefore, in some scenarios, an interpolation search can provide an accurate estimation of the closest solution to the search problem. The ternary search method is a variant of the binary search algorithm that has a slower temporal complexity [1, 2]. The meta-binary search algorithm is a variant of the binary search method that iteratively creates the index of the desired value within the array. The approach operates in a way akin to the binary search algorithm, exhibiting a time complexity of $O(\log n)$ for locating the desired element. The ternary search technique partitions the array into three segments, utilizing the central point of each segment to locate the desired element. The logarithmic complexity of the search algorithm is determined by the number of steps required to locate the desired element, which is logarithmically proportional to the size of the array, denoted as n . However, the search range, which represents the number of elements that need to be examined during the search, is three times the size of the array ($3n$). Consequently, the worst-case running time of the algorithm can be expressed as the logarithm of three times the size of the array ($\log 3n$), while the best-case running time may be approximated as being linearly proportional to the size of the array ($O(n)$). It is important to note that these

running time estimates are valid only when the array is sorted [3]. The jump search algorithm is a searching technique designed for sorted arrays. It involves performing a fixed number of jumps, denoted by k , on a block of data in order to locate the target element. Within each block, linear search operations are conducted to identify the desired element. The algorithm under consideration is superior to the sequential search method, but it falls short of the binary search technique. Specifically, it requires $m-1$ additional comparisons compared to sequential search, where m represents the size of the block to be traversed [4]. The interpolation search algorithm is a method that generates additional data points within a given range of known data points. Its time complexity is $O(\log \log n)$ for datasets with uniform distributions, and $O(n)$ in the worst-case scenario. The proposed approach represents an advancement over the binary search technique by employing comparison-based approaches that leverage a mathematical formula to approximate the location of the target element based on its value. Subsequently, the search is conducted in the vicinity of this estimated position. This alternative method has the potential to outperform the binary search algorithm under certain circumstances [5]. The exponential search algorithm operates on a sorted array. It begins by selecting a subarray of size 1, then doubles the size of the subarray in each iteration. The algorithm compares the final element of each subarray until the desired element is found. The algorithm in question is commonly referred to as exponential search, which exhibits a temporal complexity of $O(\log n)$ [6]. In the realm of searching for an item within an array of size n , two commonly employed strategies can be identified. The first strategy, known as sequential search, is applicable to unsorted arrays. The second technique, known as binary search, is exclusively applicable to arrays that have been sorted.

The study presented in this paper aims to introduce a novel searching algorithm that operates on a sorted array, relying solely on mathematical operations and a computed prediction approach implemented through a straightforward data structure. The search process in question exhibits a constant time complexity. While the space complexity may exceed that of sequential search, the time complexity can be lowered to $O(1)$ in numerous scenarios and to $O(\text{constant})$ in the worst-case scenario. The algorithm relies on the computation of the array's average, operating under the assumption of a uniform distribution of items within the list. Additionally, it generates supplementary arrays, one of which records the frequency of successful matches, while the other stores the locations where the sought-after items can be located. Part of process is similar to the Knuth-Pratt-Morris algorithm, which is commonly used for pattern matching [7]. The approach under consideration aims to decrease the time complexity by minimizing the number of comparisons and implementing a straightforward prediction system that ensures the absence of collisions through the utilization of error-free lookup tables and basic arithmetic operations.

The primary objectives underlying this research endeavor are to provide a straightforward predictive approach utilizing search techniques and to execute a basic computation with constant time complexity.

The subsequent sections of this study are structured as follows: Section II presents a review of relevant literature pertaining to searching algorithms. It is worth noting that the process of locating sufficient recent works on searching algorithms proved to be challenging. The majority of the literature discovered consisted of dated publications or encompassed broader discussions on binary search algorithms. In Section 3, the proposed algorithm is introduced. In Section IV are shown the results obtained from the proposed study, followed by an examination of the algorithm employed. Finally, the paper concludes with findings and discusses potential avenues for further research in Section V.

II. RELATED WORKS

One of the main benefits of searching operations in computer science is their capability to assist in finding particular data from databases. The effectiveness of the search process directly impacts the overall performance of the system. Searching algorithms are widely employed in several computer applications, such as problem-solving [9], data analysis, and information retrieval, due to their ability to efficiently search through extensive datasets within a limited timeframe. Various searching techniques exist, including sequential, binary, hashing, and graph search. The selection of an appropriate algorithm is contingent upon the particular situation at hand and the attributes of the data being queried [10]. The authors of [5] introduce a hybrid search method known as interpolated binary search (IBS), which integrates the interpolation algorithm and the binary search algorithm to accurately determine the precise position of the desired object. The IBS algorithm commences by employing an interpolation algorithm to estimate the approximate location of the item being searched. It subsequently operates as a binary search algorithm to precisely determine the location of the sought-after item. The Inverse Binary Search (IBS) algorithm exhibits a greater computational cost compared to both the binary search algorithm and the interpolation technique. However, when executed on uniformly distributed data, IBS demonstrates a lower temporal complexity cost than the aforementioned algorithms. Specifically, its time complexity is $O(\log 2 \log 2n)$. On the other hand, when applied to non-uniformly distributed datasets, IBS necessitates $O(\log 2n)$ operations. In [8], a comparison between different search algorithms is presented, the authors analyze the performance of various search algorithms, including uninformed search algorithms (DFS, uniform cost search) and informed search algorithms (A^* and BFS), with a focus on their time complexity and space complexity.

Graph search algorithms typically construct a graph based on the given input data and traverse the nodes of the graph using various strategies in order to locate the desired objects [13]. The algorithms that were examined all exhibited a time complexity of $O(mb)$, where m represents the number of offspring for each node (also known as the branching factor) and b represents the solution depth, which is the length of the path. The binary search algorithm utilizes a value of m equal to 2 and assigns b as the logarithm base 2 of n .

The act of searching is a crucial and widely employed process in several contexts. In order to obtain data of various

types, it is necessary to carry out two fundamental procedures: searching and sorting. In order to address this concern, it is worth noting that the majority of searching algorithms operate on an array that has been sorted. However, it is important to acknowledge that the computational expense of sorting the data can vary significantly, ranging from a best-case scenario of $O(n \log n)$ to a worst-case scenario of $O(n^2)$. This additional cost must be taken into consideration when evaluating the overall efficiency of the search process. The authors of [11] present a novel technique called the bound sequential search (BSS) algorithm, which uses logical gates to simultaneously search for two items. The primary concept is around the utilization of Binary Search with Sorted Subarrays (BSS) to concurrently search for about two items, hence eliminating the need for executing the search process twice, all without the involvement of parallel processors. In the context of Binary Search Systems (BSS), the process of looking for two items, X and Y, involves the utilization of an additional key, Z. This key, denoted as Z, is determined by the logical operation of the inclusive OR between X and Y. Subsequently, Z serves as the key element for conducting a search operation within an unsorted array, denoted as A, which possesses a size of n. The search procedure commences by evaluating each element in the array A against Z, provided that $Z \text{ AND } A[i] \neq A[i]$ indicates that neither of the two elements is present in A[i].

In the event of the most unfavorable scenario, the computational complexity for locating two items in the BSS (Binary Search Structure) is N, as opposed to $2N$ in sequential search or $(n+2 \log n)$ in the binary search algorithm. The Bubble Sort algorithm demonstrates superior performance in terms of comparison count in both the best and worst case scenarios, when compared to the Binary Search algorithm and the Sequential Search method. A limitation of this algorithm is its inability to perform a search for a single item, as it is designed to search for about two objects simultaneously.

Hashing is a technique employed in the search of extensive databases, wherein a distinct key is generated for each record or item. This key serves to designate the specific area within the database where the item is potentially kept. The time complexity for retrieving an item from a big database using hashing can range from constant time ($O(1)$) to linear time ($O(n)$) in the worst-case scenario. The variable n represents the size of the linked list, which is utilized for the purpose of storing colliding data within a single slot, denoted as [12]. The objective is to verify the presence of item X inside a given dataset. As the volume of data expands, the intricacy of the situation escalates. The problem of searching has been introduced and elucidated by Levin and Solomonoff throughout the time span of 1973 to 1984 [14,15, 16, and 17]. In reference [18], a novel quantum technique is presented for the search problem, demonstrating a polynomial time complexity. This algorithm utilizes XOR logical gates to convert the data into a polynomial form, following the amplification process provided in reference [19]. Consequently, the search operation may be executed within a polynomial time frame relative to the input size, denoted as n.

In the cited work [20, 21], the authors introduce a fractional cascade algorithm, which is a method aimed at

enhancing the efficiency of binary search algorithms. This methodology achieves a reduction in the time complexity of binary search algorithms to $O(k + \log n)$ while searching for k elements within a sorted array of size n.

III. HUSSEIN SEARCH ALGORITHM

The Hussein search algorithm is designed to efficiently locate an element in a sorted array. It achieves this by utilizing a prediction table structure and a sequential search algorithm. This approach reduces the time complexity from $O(n)$ in sequential search or $O(\log n)$ in binary search to $O(1)$ [7]. The algorithm achieves this improvement by employing only arithmetic operations and a technique inspired by the interpolation method for searching sorted lists [1]. It is important to note that the algorithm assumes a uniform distribution of elements in the list. The algorithm operates in two distinct phases: the preprocessing phase and the searching phase. During the preparation step, the entirety of the array undergoes arithmetic operations in order to ready the data for the subsequent prediction finding procedure. During the preprocessing phase, an array of integers denoted as A, which has a size of n is examined. The data within this array is created in a random manner. During this phase, the operations conducted involve the calculation of the average value of variable A. The average can be determined by calculating the mean of a set of values. To obtain the mean, each element in array A should be divided by a given value t, and the resulting values should be stored in a new array B. The size of array B is denoted as n. Generate a novel array C, whose size is determined by the floor function applied to the value of t. To iterate through a set of counters starting from 0 to size (c), the ceiling of each element $B[i]$ is compared with the indices of C. The number of matching values is determined, and the resulting matching score is stored in a new list C1 at the corresponding index of the matching item. The elements contained within the array C1 will represent the respective indices of the desired item within the original array. During the searching phase, the following operations are performed to determine whether an item X is present in an array: X is divided by t, and the resulting value is compared with the indices of C1 using the ceiling function. The value found in the corresponding cell of C1 represents the index of the searched item X in the original array. If the corresponding cell is empty, it indicates that the item X is not present in the array.

The following example provides a comprehensive elucidation of the functioning of the Hussein search method. In this scenario, array A of size 16 is randomly produced by [7], and the data within the array is uniformly dispersed. The elements in set A are arranged in ascending order, while the outcomes of dividing each element in set A by the average of the average are recorded in set B (refer to Fig. 1(a)). A new array, denoted as C, is to be created with a size of 30. This array will contain the floor value of the average values obtained from Fig. 1(b). The quantity of corresponding elements in the array is stored in Fig. 1(c). Now, let us explore the scenario where aim to search for item X, which has a value of 310, within the table. Initially, the value of X is divided by 30.89. The resulting quotient is then rounded up to the nearest whole number, denoted as 11. Subsequently, proceed to locate the element within the array by referring to the index position

11. Upon examination, it is determined that the value at this index is 0, indicating that the desired element is not present inside the array. Let X be equal to 275. Next, perform a division operation on X by t, and subsequently apply the ceiling function. The resulting value, 9, is assigned to the index 9 in array C. Upon further examination, we finally, observe that the element at index 9 in array C is 1, indicating that X is located at index 5 in the original array (as depicted in Fig. 2).

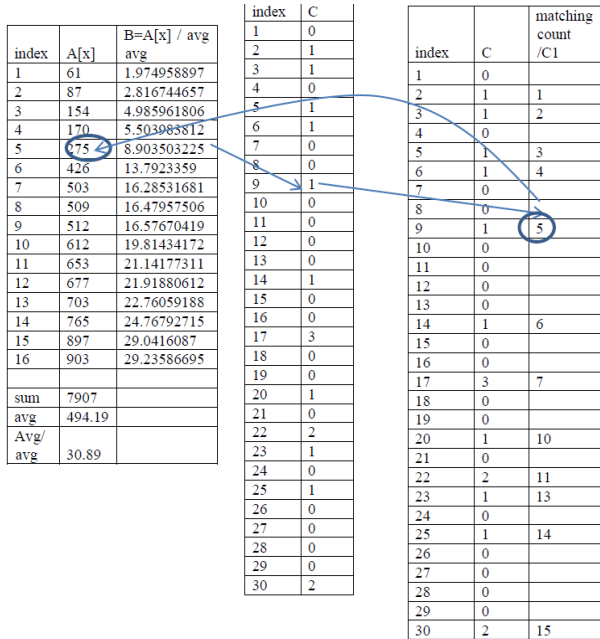


Figure 1a: the preprocessing phase figure 1b: index matching figure 1c: the matching score

Fig. 1. Hussein search algorithm simulation on random data generation.

5 random tests					
X	/ avg avg	ceiling	from	to	count
275	8.90255746	9	5	5	1 exist
250	8.09323406	9	5	5	1 no exist
310	10.0356102	11	/	/	0 no exist
900	29.1356426	30	15	16	2 no exist
1003	32.470055	33	/	/	/ no exist

Fig. 2. Testing the search operation using Hussein search algorithm.

A. Algorithm Description:

1) *Preprocessing phase:* Consider we have an array of integers A of size n, where the data is randomly generated, in this phase the following operations are performed:

- Calculate the average of A. $average = \frac{sum(A)}{n}$
- Calculate the average of average: $t = \frac{average}{n}$
- Divide all the elements of A by t and store it in a new array B, the size of B is n
- Create a new prediction array C with size floor (t)
- For I counter that start from 0 to size (c), compare the ceiling of B[i] with the indices of C, and count the

matching values and store the matching score in a new list C1 in the corresponding index to the matching item. The values stored in the array C1 will be the corresponding index of the searched item in the original array.

2) *Searching phase:* For searching an item X, if in the array or not is performed the following operations:

- Divide X by t, $S = \frac{X}{t}$
- Compare the ceiling of (S) with the indices of C1; the value found in the corresponding cell will be the index of the searched item (X) in the original array, if the corresponding cell is empty that means the item is not found in the array.

IV. RESULTS AND ANALYSIS

The proposed algorithm is simple and easy to understand and implement. It is implemented and tested in MacBook air i5 processor 1.3 GHz speed, 8 GB ram, using visual studio, C# and in Python, and is tested for large input size list up to 16 mg . Furthermore, we have successfully implemented both the binary search method and the ternary algorithm using the C# programming language. These implementations were carried out in an identical setting, with the input size being consistent with that of the Hussein search algorithm. The objective of the experiment was to conduct N iterations in order to seek a randomly produced list with a size of N. The findings indicate that the speed of the prediction searching method in all evaluated algorithms exhibits a linear relationship with the amount of input. However, it is noteworthy that as the input size increases, the performance of the Hussein search surpasses that of the other algorithms. The Hussein search method exhibits a constant time complexity of O(1) for finding an individual item.

Consequently, the search operation for n items may be accomplished in linear time complexity of O(n). This stands in contrast to the binary search strategy, which necessitates a time complexity of O(n log n) for searching n items. Fig. 3 presents the outcomes achieved by the Hussein search algorithm in contrast to the other algorithms when searching for N items across varying input sizes. while considering input sizes of 8 MB and 16 MB, it has been observed that the Hussein search method exhibits a time requirement that is 20% lower than that of the binary search strategy, and 17.3% lower than that of the ternary algorithm, while searching for all items. Hussein's search method demonstrates a search speed that is approximately 494% greater than that of binary search when applied to a dataset of 16 MB. Table 1 illustrates the speedup, which quantifies the extent to which the Hussein search algorithm outperforms the binary search algorithm across various input sizes. Table I illustrates the observed increase in search speed across various input sizes.

In the Hussein search algorithm the searching process about an item requires O (1), which means searching n items requires only O(n) in comparison with the binary search algorithm that requires O(n log n). Fig. 3, show the results obtained by the Hussein search algorithm in comparison to the other algorithms for different input size. For input size (8 M,

16 M) searching all the items using the Hussein search algorithm requires time that is 20% smaller than the binary search algorithm and 17.3 % smaller than the Ternary algorithm. The search speed in the Hussein search algorithm is increased by about 494% than the binary search for 16 M of data. Table I, and Fig. 3 show the speed up that represents how much Hussein's search is faster than the binary search algorithm in searching about all the input sizes.

TABLE I. THE SPEED UP IN SEARCHING DIFFERENT INPUT SIZE

	Binary	Trenary	Hussein	Speedup
4 k	2	2	4	50%
8k	5	6	2	250%
16 k	4	6	2	200%
32 k	12	13	3	400%
64 k	27	24	5	540%
128 k	43	22	10	430%
256 k	89	109	20	445%
512 k	189	222	39	485%
1 MB	349	419	69	506%
2MB	703	825	140	502%
4MB	1398	1606	259	540%
8MB	3343	3568	629	531%
16MB	5813	6784	1176	494%

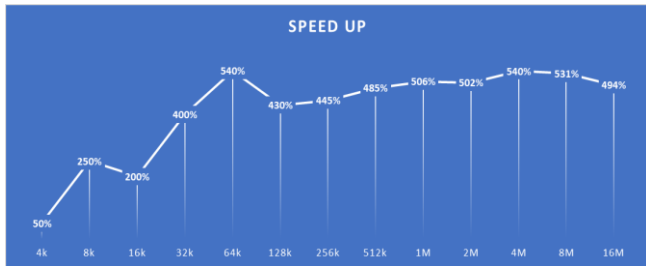


Fig. 3. Hussein search algorithm speed up.

```

static float prehusn1(int[] arr)
{
    float avg;
    avg = (arr[0]/2) + (arr[1]/2);
    for(uint i=2;i<arr.Length;i++)
    {
        avg = (avg * i / (i + 1)) + (arr[i] / (i + 1));
    }
    return avg/arr.Length;
}

static int prehsn2(float avg, int[] arra, uint[] arrb, uint[] arcc)
{
    for(uint i = 0; i<arra.Length;i++)
    {
        uint l = (uint)(arra[i] / avg);
        arrb[l]++;
        if (arrb[l] == 1)
            arcc[l] = i;
    }
    return 0;
}

```

Fig. 4. The implementation of mathematical operations in Hussein search algorithm.

Implementation of Hussein Search Algorithm: as mentioned previously Hussein Search algorithm is implemented in C#, with binary search algorithm and ternary algorithm on the same environment and the same data. Fig. 4 and 5 represent the implementation of the main functions of the Hussein search algorithm.

```

static int HusnSearch(int x, float avg, int[] arra, uint[] arrb, uint[] arcc)
{
    uint l = (uint)(x / avg);
    if (l<arrb.Length)
    if (arrb[l]>0)
    {
        uint m = arcc[l];
        for(uint i =m; i<m+arrb[l];i++)
        {
            if (arra[m] == x)
                return (int)m;
        }
        return -1;
    }
    return -1;
}

```

Fig. 5. The implementation of the Hussein search algorithm.

V. CONCLUSION AND FUTURE WORKS

In this study, we introduce the Hussein search algorithm, a novel informed search approach that leverages a straightforward prediction method, basic arithmetic operations, and a simple data structure. The findings demonstrate that the Hussein search algorithm outperforms previous search algorithms in terms of time complexity, particularly when dealing with substantial data sets. Moving forward, there are several avenues for future research. Firstly, it would be valuable to explore the algorithm's performance under different search scenarios and input distributions. Additionally, investigating potential optimizations and further enhancements to the algorithm could yield even more efficient search capabilities. Finally, conducting comparative studies with other state-of-the-art search algorithms would provide a comprehensive evaluation of the Hussein search algorithm's effectiveness. The procedure operates on an array that has been sorted, with the underlying assumption that the data is spread equally. The Hussein search algorithm has a time complexity of $O(n)$ for finding n items. In contrast, the binary search algorithm has a time complexity of $O(n \log n)$, while the sequential search technique requires $O(n^2)$ in the worst case. Given the assumption of a sorted array, the proposed technique offers a notable advantage in terms of computational simplicity and a reduction in the number of comparisons required for item search. Specifically, the algorithm achieves a time complexity of $O(1)$ when searching for an item by index rather than by value. Fig. 6 presents a comparison of the running times for various input sizes. The future objective is to enhance the algorithm based on simple prediction methods to operate with the same time complexity for an unsorted array. Additionally, we aim to offer a sorting algorithm that utilizes the same mathematical processes and achieves linear time complexity in the worst-case scenario.

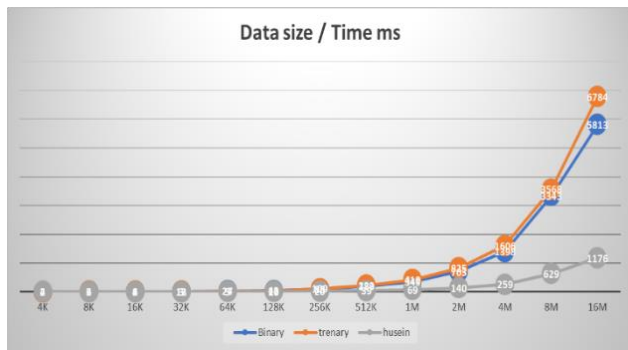


Fig. 6. Running time comparison for different input size.

REFERENCES

- [1] C. Blum, A. Roli: Meta Heuristic in combinatorial Optimization: overview and Conceptual Comparison. ACM Computing Survey, vol 35, No.3, PP: 268-308.2003.
- [2] Brodrick Crwford, Ricardo So et al: Putting Continuous Meta Heuristic to Work in Binary Search Spaces. Hindawi complexity. Volume 2017. <https://doi.org/10.1155/2017/8404231>. PP:1-19.
- [3] Manpreet Singh Bajwa, Arun Prakash Agarwal, Sumati Manchanda. Ternary Search Algorithm: Improvement of Binary Search. 2nd International conference on Computing for Sustainable Global Development (INDIA COM)11-13 March 2015. New Delhi. India. ieeexplore.org/document/7100542.
- [4] Adnan Saher Mohammed, Sahin Emrah Amrahou, Fatin V. Celebi. Interpolated Binary Search : An Efficient Hybrid Search Algorithm on Ordered Datasets. Engineering Science and Technology, An International Journal. Volume 24, Issue.5, October 2021, PP:1072-1079. <https://doi.org/10.1016/j.jestch.2021.02.009>.
- [5] Milos Simic. Exponential Search. Last modified 26 June 2022 available at: <https://www.baeldung.com/cs/exponential-search>.
- [6] Knuth, Donald (1998). Sorting and Searching: The art of Computer Programming. (2nd ed).MA. Addison- Wesley. ISBN:978-0-201-89685-5.
- [7] Maharshi, J. Pthak, Romit L. Patel, Sonal P. Rmi. Comparative Analysis of Search Algorithm. International Journal of Computer Applications. Volume 179. No.50. June 2018. PP:40-43.

- [9] Dhanya Thailappan. Introduction to Problem Solving Using Search Algorithms for Beginners. Last modified 26th July 2022. Analytics Vidhya. Available at: <https://www.AnalyticsVidhya.com/blog/2021/an-introduction-to-problem-solving-using-search-algorithms-for-beginners>.
- [10] Stuart Russell, Peter Norvig. Artificial Intelligence: A Modern Approach. Pearson Education International. 4th Edition 2021. ISBN-13: 978-1292401133.
- [11] Omer H. Abu El Hija, Azmi Alazzam. Bound Sequential Search (BSS). Proceedings of the Worlds Congress on engineering and Computer Science. 2012. Vol.1, WCES 2012. October 24-26/2012. San Francisco, USA.
- [12] George T. Heineman, Gary Pllice, Stanley Sekow. Algorithms in a Nuthshell: A Practical Guide. 2nd Edition O'reilly Media (April, 12, 2016). ISBN-10: 1491948922.
- [13] Thomas Cormen, Charles E. Leiserson, and Ronald L. Rivest. Introduction to Algorithms. 4th Edition (2022). The MIT press. ISBN-10: 026204630X.
- [14] M. R Gavey and D. S. Johnson. Computer and Intractability: A Guide to the Theory of NP-Completeness. W. H. freeman, 1979.
- [15] L. A. Levin. Universal Sequential Search Problems. Problems of Information Transmission. 9(3): 265-266-1973.
- [16] L. A. Levin. Randomness Conversation Inequalities. Information and Independence in Mathematical Theories. Information and Control. 61: 15-37. 1984
- [17] R. J. Salmonoff. Optimum Sequential Search. Memorandum, Oxbridge Research Cambridge, Mass. June 1984.
- [18] S. Iriyama, M. Ohya, and V. Velovich. On Quantum Algorithm for Binary Search and its Computational Complexity. arXiv:1306.5039v1 [quant-ph] 21 June 2013.
- [19] M. ohya and I. V. Vovich. Quantum Computing and Chaotic amplifier. J.OPT.B, 5, No.6. 639-642.2003.
- [20] Chazelle, Bernard, Liu, Ding. Lower Bounds for Intersection Searching and Fractional Cascading in Higher Dimension.33rd ACM Symposium on Theory of Computing. ACM. PP 322-329. Doi: 10:1145/380752.380818. ISBN:978-1-58113-349-3. Retrieved 30 June 2018.
- [21] Chazelle Bernard, Liu, Ding (1 March 2004). Lower Bound for Intersection Searching and Fractional Cascading in Higher Dimension. Journal of Computer and System Sciences. 68(2): 269-284. Doi:10.1016/j-jcs.2003.07.003. ISSN 0022-0000. Retrieved 30 June 2018.

Testing the Usability of Serious Game for Low Vision Children

Nurul Izzah Othman, Hazura Mohamed, Nor Azan Mat Zin

Faculty of Information Science and Technology, The National University of Malaysia, 43600 Bangi, Malaysia

Abstract—Serious games are prodigious tools for building language, science and math knowledge and skills. Despite a growing number of studies on using serious games for learning, children with visual impairment have obstacles when playing the games. Low vision children have a visual balance that can be assisted with assistive technology. A 2D serious game for learning Mathematics is developed using Unity for low vision children. In order to enhance the game's accessibility for low vision children, accessibility elements have been implemented in the serious game prototype. Those elements are screen design (buttons, menus, and navigation), multimedia (text, graphics, audio, and animation), object motion, and language. Upon completion of the serious game, usability testing was done to identify the accessibility of the serious game to low vision children based on the usability level. The observation technique is used for analysing the serious game. The overall usability score is good based on aspects of effectiveness, efficiency and user satisfaction tested.

Keywords—Serious game; learning; low vision; usability; accessibility

I. INTRODUCTION

Nowadays, information and communication technologies (ICT) are used in education. We can find a lot of educational applications that can help young learners learn. Several ICT applications are on the market, such as e-books, multimedia courseware, and games [1] [2]. Gaming is a very popular activity enjoyed by children. Thus, a serious game is the most promising tool for helping children learn. A serious game is used beyond entertainment [3][4]. Serious games are prodigious tools for building language, science, and math knowledge and skills. Despite a growing number of studies on using serious games for learning, children with visual impairments have obstacles when playing the games. There are two categories of visual impairment: blindness and low vision. Low vision children have a visual balance that can be assisted with assistive technology. The World Health Organisation defines low vision as visual acuity between 20/70 and 20/400 with a possible correction of 20 degrees or less [5]. Low vision children have obstacles accessing graphical elements in games. A preliminary study conducted by [6] indicates that low vision has several accessibility issues, such as visual, animation, audio, and navigation. They need to look at the graphical elements from a very close distance because the size of the text and graphics is small. The colour contrast between the graphics and background is low. The choice of dark colour is unsuitable for the children. They also have difficulties navigating the menus and buttons in games. They also focus on one sound at a time. Besides, fast animation movements affect their vision

when playing games. Thus, to ensure they can play the game, a 2D serious game for learning Mathematics is developed using Unity for low vision children. This serious game runs on mobile devices using the Android operating system. This game's storyline is about a rabbit named "Bunny" trying to save his friends who a tiger kidnaped. Bunny needs to complete the Mathematical tasks provided for each game level to obtain the instructions and tools used in the next challenge. Hints and tools will help Bunny save his friends from the tiger. This game consists of three levels of Mathematical tasks, which are based on the Mathematics Syllabus for Year 1. The game's content consists of an introduction to numbers, shapes, addition, and subtraction.

Serious games should be easily accessible to low vision children. The game design should be flexible, with an interface adapted to the accessibility requirements of low vision children. In order to enhance the game's accessibility for low vision children, accessibility elements have been implemented in the serious game prototype. Those elements are screen design (buttons, menus, and navigation), multimedia (text, graphics, audio, and animation), object motion, and language. The children can set the appropriate background colour and contrast level to see menus and buttons more clearly based on their vision level. The game's navigation is designed to be consistent throughout the game. The text size is large and adapted to the children's vision. The background colour has a high colour contrast. The use of bright colours and large graphics are implemented in the game. Children can adjust the background colour. Thus, children can easily identify and move objects in the game. The background audio must be clear and adjustable based on the needs of children with low vision. Important objects in the game include background sounds to help the children identify the position of the object. The game's task instructions are accompanied by background audio so children can easily understand the game's storyline. Besides, the language used in the game should be easy to understand. For the proposed serious game, the Malay Language is used because it is used in learning Mathematics in primary school. The movement of objects in the game can also be adjusted according to the child's vision so that it is not too fast. The movement of objects in the game is also minimal so that children can control the game based on their needs. Upon completion of the serious game, usability testing was done to identify the accessibility of the serious game to low vision children based on the usability level. Usability testing is conducted to ensure the serious game fulfills the children's accessibility requirements. The serious game user interface is depicted in Fig. 1.

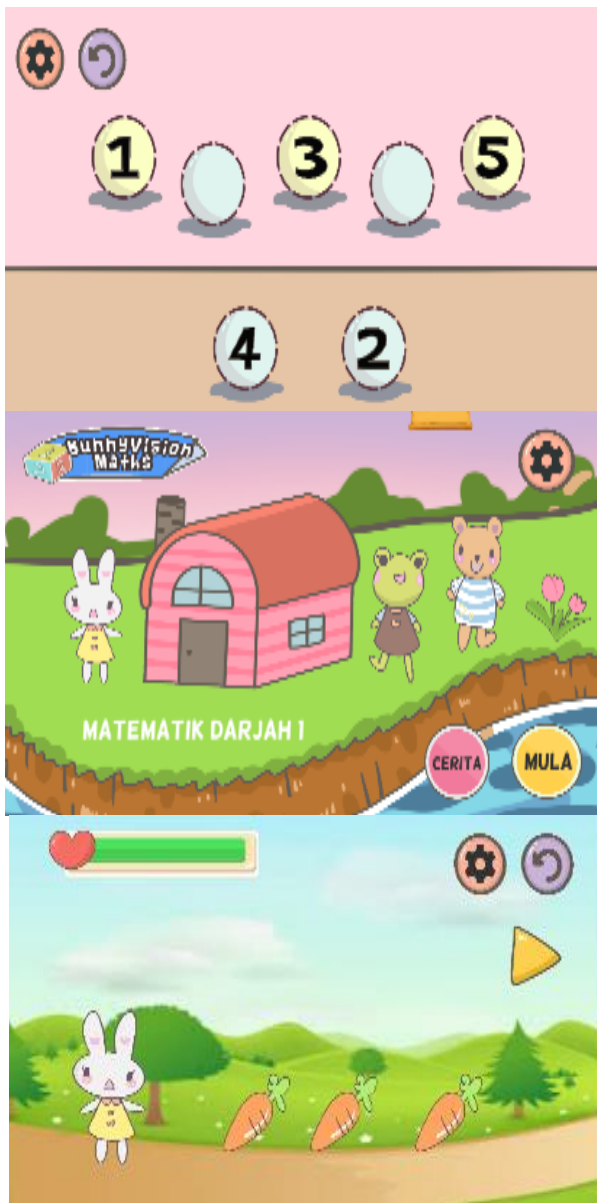


Fig. 1. Serious game user interface.

Thus, this paper presents the usability testing of Math Game, a serious game prototype that was developed specifically for low vision children aged seven years old. The serious game is developed based on the accessibility requirements of low vision children. The effectiveness, efficiency, and user satisfaction of the serious game are tested in the usability testing. The observation technique is used for analysing the serious game. The usability testing aims to identify the serious game's usability level for low vision children.

This paper is organised as follows: Section I discusses introduction, while Section II describes the related work. Section III describes the methodology and Section IV and Section V discusses the results and discussion, respectively, while Section VI presents the conclusion and plan for future work.

II. RELATED WORK

The International Organization for Standardization (ISO) 9241-11 (1998) defines usability as the ability of a product to be used by users to accomplish certain goals effectively, efficiently, and with user satisfaction [1]. Usability also refers to whether a product is easy to use. Usability testing allows a product to be usable by target users to meet accessibility requirements. Based on the ISO 9241-11 definition, usability consists of the following components:

Effectiveness: The accuracy and completeness with which a user can achieve a certain goal in a certain environment.

Efficiency: The effort the user gives to complete a task and achieve the objective of using an application or system.

User Satisfaction: User comfort and acceptance of the system.

Usability testing is conducted to identify usability problems before releasing the serious game in its real context. The usability testing ensures the serious game fulfills the children's accessibility requirements. It is because both usability and accessibility should be considered good design practices and involved in system and application development stages. There is some overlap between the concepts of usability and accessibility. Accessibility enables particular users to access ICT applications independently without accounting for their disability. At the same time, usability refers to the ability of the ICT application to carry out the intended function effectively, efficiently, and with satisfaction when used by users. Therefore, accessibility specifically refers to disabled users, while usability refers to the general population. According to [7], usability testing is also suitable for assessing ICT applications for accessibility.

In human-computer interaction (HCI), usability is one of the focus areas where specified users can use a product to achieve goals effectively, efficiently, and satisfactorily [8]. When designing for usability and accessibility, a usability study should be conducted on the ICT application. Usability is defined as how efficiently user requirements are fulfilled [9]. It has developed usability criteria for accessible websites based on their effectiveness, where the user carries out tasks without experiencing barriers such as clear link texts. For efficiency, the user obtains the desired information quickly, and the system is usable. In contrast, for satisfaction, the user feels joy when navigating the website because the screen design facilitates navigation by the disabled user and avoids the accessibility barrier.

Several evaluation factors in usability testing include screen appearance, consistency, accessibility, navigation, interactivity, and content [8]. The literature comprises many usability testing types, methods, measurements, and respondents. Besides, usability research could be done in many forms based on the scope and goal of the ICT application. There are activities for usability research, such as focus groups, task analysis, user observation, interviews, and surveys. The activities provide insights into how users interact with the ICT application [10]. For example, from the observation and interview, the respondent's reaction when using the product will help the researcher understand the user's satisfaction with the ICT

application [11]. A summative usability evaluation has been done using observation and questionnaires to measure the usability of the Mudah.my mobile application's efficiency, effectiveness, and satisfaction [12]. Ten respondents are involved in the study and will perform five application tasks. Based on the study, the results show that the Mudah.my mobile application is easy to use and consistent. A usability study was also done using the Technology Acceptance Model to evaluate the usability of virtual game-based simulation to help nursing students improve their pediatric nursing skills [13]. The methods involved heuristic evaluation with experts, think-aloud activities while playing the game, and interviews. The study shows high user satisfaction, and they learned about pediatrics care easily. Observation has also been conducted to evaluate the usability of the TVET m-learning application for 30 Multimedia Software Technology course students. A questionnaire was used to measure the usability elements such as system usefulness, ease of use, ease of learning, and user satisfaction. A special room and tools, such as a video camera, notebooks, mobile devices and task lists, are prepared for this usability testing. This study's results showed that students accepted the TVET m-learning application, and the usability score value was at a high level [14].

In addition, usability metrics have been implemented to evaluate the quality of the ICT applications. However, the metrics used for usability testing always change due to new inventions in ICT applications [15]. There are numerous usability measurement tools for usability testing. Website Analysis and Measurement Inventory (WAMMI) is a usability measurement tool that assesses website usability. WAMMI is a questionnaire developed by the Human Factors Research Group (HFRG). WAMMI comprises usability factors such as attractiveness, controllability, efficiency, and helpfulness [16]. WAMMI is used to evaluate the usability level of the Ministry of Education Malaysia (MOE) web portal and provide usability enhancement based on the testing results. There are two stages of usability testing, such as pre-usability testing to evaluate the current web portal and post-usability testing for the web portal that has been enhanced. The pre-usability testing result showed that the usability of the current web portal was at a moderate level. In contrast, the post-usability testing result showed that usability improved significantly [16].

Besides, the Software Usability Measurement Inventory (SUMI) was used as a usability measurement tool to measure system usability. SUMI was used to evaluate user experience when it was introduced in the 1990s. SUMI consists of usability criteria such as effectiveness, efficiency, satisfaction, error management, consistency, adaptability, and compatibility. This evaluation tool assesses a semantic image retrieval application, WebSIR. The results showed that 85% of respondents were satisfied with this application, which is easy to use [16]. Another study by [17] evaluated the usability of South Tangerang E-Government using SUMI. Based on the study, the usability scores of the e-government website are taken as a benchmark for the system's usability. The system usability scores are 85 for effectiveness, 81.5 for efficiency, and 72.5 for satisfaction. Thus, the usability of the e-government website is good.

A representative sample of users must conduct usability testing because users are different and have different problems [18]. Suitable usability testing methods and measurement should be considered when planning the testing with users, especially children. Usability testing with children consists of introspection, direct observation, thinking aloud, and interaction. Simple observation is used to observe users as they perform their tasks. Thus, this technique has been recognised as one of the best techniques for usability testing [19]. The usability testing should be planned before it is conducted on a real user.

III. METHODOLOGY

The usability testing method used in this study is direct observation. The observation method is conducted to evaluate the user experience of the product. This method involved evaluating users by observing them carrying out the tasks provided. Observations are conducted based on effectiveness, efficiency, and user satisfaction. User satisfaction is subjective. Thus, interviews were also conducted with low vision children to get feedback regarding their satisfaction level with the tested serious game. During the interview, emoticons (like, normal, dislike) are prepared to observe the user's reaction to the game. Observations will be recorded on the user satisfaction checklist. Users are required to perform tasks using the provided prototypes.

This usability testing used a selected study sample conducted at Sekolah Kebangsaan Pendidikan Khas Jalan Batu, Kuala Lumpur, and Sekolah Kebangsaan Pendidikan Khas Muar. This testing involves fifteen to seven-year-old low vision children (nine boys and six girls). In order to conduct usability testing, there are ethics that the researcher must follow. Since the testing was conducted during the COVID-19 pandemic, it complied with the Standard Operating Procedures (SOP) set by the Ministry of Education. Before conducting the usability test, consent from parents or guardians is required to ensure that the study meets research guidelines and ethics. The purpose of this research is also explained to the teachers involved.

A. Usability Testing Procedures

The activities involved in usability testing include determining the testing objective, preparing the instruments and testing tools, identifying the task scenario, preparing the testing checklist, and measuring usability testing. The testing procedure begins with the preparation of the testing instruments and tools. The testing procedure is further explained in the following sections.

1) *Determination of testing objective:* This usability testing was conducted to verify the accessibility of serious games to low vision children. Evaluation is important to ensure serious games accomplish development objectives. The evaluation is conducted based on the ISO 9241-11 definition, which covers effectiveness, efficiency, and user satisfaction concerning serious games.

2) *Testing instruments and testing tools preparation:* The testing instruments are prepared before the test is conducted. The testing instruments are reviewed by lecturers who are

experts in multimedia and games. The testing instruments consist of a serious game prototype, a usability testing task list, an observation checklist, a video recording, and a screen recording. Testing tools, such as tablets, a video camera, and a screen recorder, are prepared to support the usability test.

3) *Identify task scenarios*: Usability testing is conducted by giving the user a task scenario while the user plays the game. Researchers help users play the game when needed. The task scenario is an action that needs to be performed by the user during interface testing. The task scenario is based on the game task instructions. The task scenario is prepared according to the game screen and the game's design, such as the start screen, the narration screen, and the game task screen. Researchers use these task scenarios to ensure that all testing activities are going as planned.

4) *Testing checklist preparation*: A checklist is prepared based on the usability construct [1], [19], which consists of effectiveness, efficiency, and user satisfaction. The usability checklist consists of sixty-three items. Checklists are provided based on game functions and actions that players need to perform while playing the game. Table I shows the contents of the usability checklist.

TABLE I. THE CONTENTS OF THE USABILITY CHECKLIST

Accessibility Elements	Effectiveness	Efficiency	User Satisfaction
Menu and Button Screen Design) and Navigation	Children can click the menu and button functions.	Children easily click on menus and buttons. Children click on the right menu and button.	Children's reactions when they click the menu and button.
Multimedia (Graphics, Animation, Text, and Audio)	The multimedia on the screen are interesting for children. The audio used appeals to children.	Children click objects in the game easily. Children do not confuse while playing.	Children's reactions to multimedia in the game.
Object Movement	Adjustable speed of objects helps children to play.	Object movement makes it easier for children to click on the correct object in the game.	Children's reactions to the object's movement.
Language	The language used in the game is understandable and suitable for the children's ability.	The language used does not confuse or cause children to make mistakes while playing.	Children's reactions to the language used.

5) *Measurement of usability testing*: The measurement of effectiveness, efficiency, and user satisfaction are measured using the usability score [15], [20]. There are three observation options, effectiveness, efficiency, and user satisfaction, for the checklist, with the score represented for each option. Scores are awarded based on completing each

task. Scores are then accumulated based on the three observation options. The usability measurements are as follows:

a) *Effectiveness Measurement*: Effectiveness measures is the ability of the user to complete a task within the game. If a task is completed successfully, a score of 5.0 will be marked on the effectiveness checklist item. Half of the task is completed or fails to complete; a score of 2.5 and 0 will be marked on the checklist item, respectively. Table II shows the scoring guide for the effectiveness checklist item during the observation.

TABLE II. EFFECTIVENESS SCORING GUIDELINES

Score	5.0	2.5	0
Details	<p>When the children manage to complete the task successfully, the accessibility elements on-screen help children playing the game as below:</p> <ul style="list-style-type: none"> -Children have no problem navigating the button and menu display on the screen. -Children understand the instructions of the game. 	<p>Children completed half of the task but had several issues, as below:</p> <ul style="list-style-type: none"> -Children have difficulty navigating certain buttons and menu displays on the game screen, but they can still proceed with the game. -Multimedia elements do not consistently attract children's attention. 	<p>Children do not complete the task below:</p> <ul style="list-style-type: none"> -Children have a problem finding all the menus and button displays on the game screen. -Children do not understand the game instructions.

Thus, the usability score for the effectiveness, X, is formulated as:

$$X = (\sum \text{Score for each effectiveness checklist item}) / \text{Total effectiveness checklist item} \quad (1)$$

b) *Efficiency measurement*: Efficiency is measured based on the duration of task completion time, which is the average task completion time. The average task completion time is a reference in efficiency testing [21]. The efficiency aspect is better when the user completes the task faster. The efficiency checklist has three observation options, with the score represented for each option. If the time is taken by the child to complete the task is equal to or faster than the average task completion time, then a score of 5.0 will be marked on the efficiency checklist item. Whereas, if the time the child takes to complete the task exceeds the average task completion time, a score of 2.5 will be marked on the efficiency checklist items. However, if the task fails, a score of 0 will be marked on the checklist items. Table III shows the scoring guide for the efficiency checklist item during the observation.

TABLE III. EFFICIENCY SCORING GUIDELINES

Score	5.0	2.5	0
Details	Children choose the right menu and button based on the game function The assistance provided when children play the game is minimal.	Some menus and buttons in the game are not selected correctly. However, they still could proceed to play the game. Some of the tasks require help.	Children are difficult to choose the right menu and button. Children play the entire game with assistance from the researcher

Thus, the usability scores for the efficiency, X, formulated as:

$$X = (\sum \text{Score for each item of efficiency checklist}) / \text{Total Efficiency Checklist item} \quad (2)$$

c) *User satisfaction measurement*: User satisfaction measurement consists of like, neutral, and disliked. A 5.0 score will be marked on the user's satisfaction checklist items if a child likes the serious game. If the child's reaction is neutral, a score of 2.5 will be marked on the checklist items, and if the child does not like the game, a score of 0 will be marked on the checklist items. Table IV shows the scoring guide for the user satisfaction checklist item during the observation.

TABLE IV. USER SATISFACTION SCORING GUIDELINES

Score	5.0	2.5	0
Details	Children show emoticons provided with "happy" signs. Children can focus on playing the game. Children finish the game successfully.	Children show emoticons provided with "neutral" signs. Children can focus on some of the games. Children stop playing in the middle of the game	Children show emoticons provided with signs "dislike". Children are not focused while playing the game. Children are unable to finish the game.

Thus, usability scores for the user satisfaction category, x, can be formulated as:

$$X = (\sum \text{Score for each item of user satisfaction checklist}) / \text{(Total User Satisfaction Checklist item)} \quad (3)$$

The average value of each usability category is considered the use score for that category. The overall usability score is the average value of usability scores for three usability categories. Thus, the overall usability score, X, can be formulated at Table V:

TABLE V. USABILITY SCORE

Weak	Moderately Good	Good
$.0 \leq x \leq 2.49$	$2.5 \leq x \leq 3.49$	$3.5 \leq x \leq 5.0$

6) *Implementation of usability test*: A pilot study is conducted on the instrument to test its reliability before it is

used in the actual testing. The pilot study involved a sample of five to seven-year-old low vision children. The pilot study was conducted to ensure that the assessment faced no problems. Through a pilot study, children's reactions, behaviours, and play tasks can be recorded. During the usability testing, children played the game with minimal assistance. The camera is used to record testing sessions. Screen recorder software records children's activities on the screen when playing the game. During the observation, researchers must observe the child's actions and reactions to the game tested. Children's actions in the game are recorded on the provided task scenario checklist and transferred to the usability checklist. Scores are given on the user's achievement in the task, which is recorded on the usability checklist based on the effectiveness, efficiency, and user satisfaction constructs. The data collected was analyzed with a descriptive analysis method. Fig. 2 shows the usability testing preparation.



Fig. 2. Usability testing preparation.

IV. RESULTS

The usability measurement is done by analysing the user's success in completing tasks. Usability data is collected as a usability score and determines whether the game design goals

are achieved and the need to improve the serious game. A user's success score indicates how successful the user is at completing a task. Scores are then accumulated based on three categories of usability: measurement of effectiveness, efficiency, and user satisfaction.

Based on the analysis, the effectiveness score for the game screen is 4.74. The children complete the tasks successfully on the game screen. Children begin to adapt in the early stages of the game. They make mistakes when selecting objects in the game and choosing the wrong answer. However, they can still reselect the correct button and complete the task. When the children can play at a higher level in the game, the user can adapt to the game interface. Fig. 3 shows the graph of the effectiveness score for each screen.

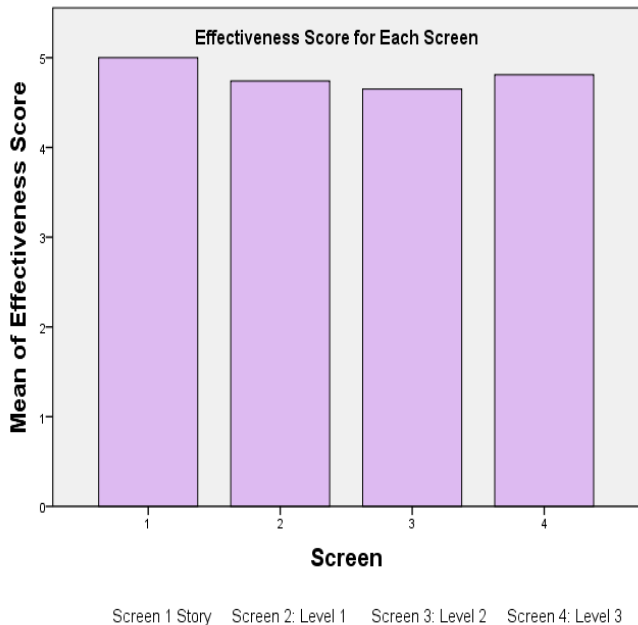


Fig. 3. The graph of effectiveness score.

Efficiency scores are given based on the time it takes to complete the task and efficiency measurement metrics. For the level 1 task, the average time the children took to complete the task was seven minutes. Children need more time to complete games because they begin to explore the game interface and begin to adapt to the game environment. The average time the children complete the level 2 and level 3 tasks is five minutes. At level 2 and 3 game tasks, children become more efficient at playing games because they can already adapt to the game interface and environment. They already understand how to play the game. The efficiency score is 4.55. Based on the usability score, the efficiency level is good. The game task can be executed smoothly and easily. Serious game efficiency analysis shows children click on the right buttons and menus to play the game. The assistance given to children is minimal, and they easily correct mistakes. Therefore, children can play the game easily and smoothly. Fig. 4 shows the graph of the efficiency score for each screen.

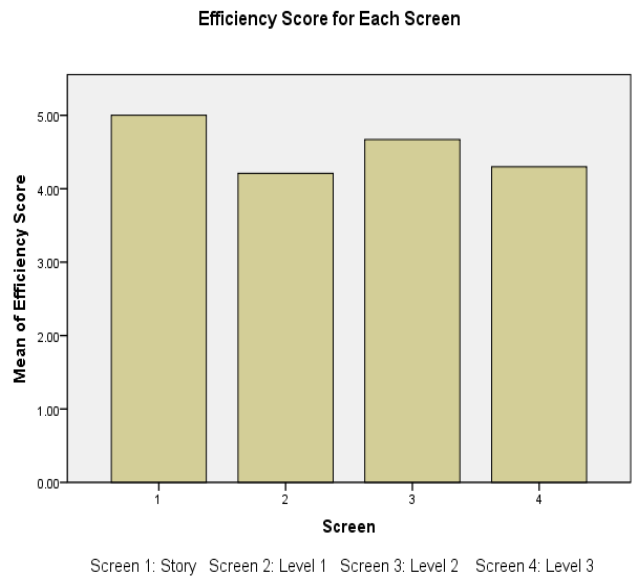


Fig. 4. The graph of efficiency score.

User satisfaction is measured based on the user's feelings when playing serious games, such as their reactions to screen design, buttons, menus, and games. For user satisfaction constructs, the user satisfaction score is 4.92. Based on usability scores, the level of user satisfaction is good. This shows that the children react positively to playing the game. Children react positively to the game interface, such as buttons, menus, graphics, and animations. Multimedia elements such as audio and in-game animations also attract users to play games. Fig. 5 shows the user the graph of the user satisfaction score.

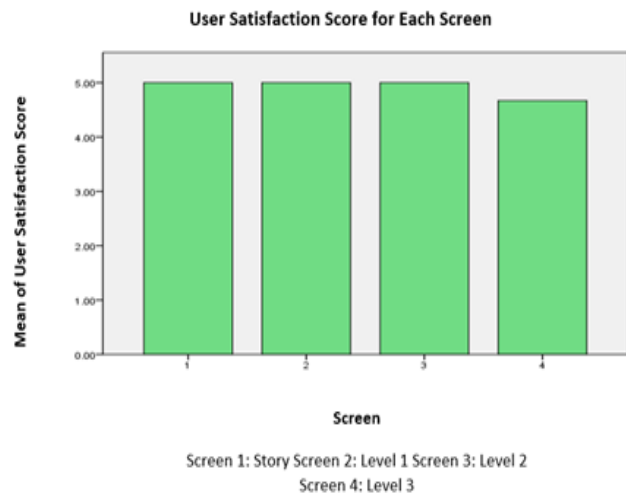


Fig. 5. The graph of user satisfaction score.

Based on aspects of effectiveness, efficiency, and user satisfaction tested, the overall usability score is 4.76, which is good. Based on descriptive usability analysis, elements of accessibility such as screen design (buttons, menus, and navigation), multimedia, language, and object motion are important in increasing the usability of serious games for low vision children.

V. DISCUSSION

In the context of the usability of serious games, accessibility elements such as screen design, multimedia, language, and object motion help children complete the game successfully. The game interface features a simple screen design with larger buttons, menus, and icons to prevent accidental clicks, enabling children to interact with the game easily. This is consistent with a previous study by Allah et al. [22], which found that screen design elements such as menus and buttons were considered for improving the usability of the user interface.

Animations provide visual and auditory feedback when objects are selected or actions are performed, aiding children to choose the correct answers during gameplay. This finding aligns with the results of Ghanouni et al. [23], who emphasized the important role of animations in game design for children. The game also includes audio to assist children with low vision in understanding the content. The option to adjust audio volume ensures that children can set it to a comfortable level or mute it if needed. This outcome demonstrates that background audio associated with objects in the game assists users in selecting the correct object or answer, as reported by Najjar et al. [24].

Brighter colors are used for important in-game objects. A good contrast between text, graphics, and background colors enhances readability, allowing children to easily identify important object. Concise language helps children understand and follow the game's tasks, and these findings are in line with previous research conducted by Benaïda [25]. However, this study goes further by demonstrating that object speed can be adjusted based on children's preferences to help children control the game.

Furthermore, when children navigate the game interface without confusion, they enjoy the gameplay and feel satisfied with the experience [26]. Multimedia elements make the game more enjoyable for children [14]. When children can easily understand the gameplay and objectives, they experience a sense of accomplishment and satisfaction. Positive reinforcement through sounds or animations can keep users engaged and motivated [27]. Therefore, the use of accessibility elements on each screen is important for helping low vision children play games.

VI. CONCLUSION

In this study, usability testing on a serious game was done successfully, involving low-vision children playing the game. The testing has been done through observation. The results of usability testing show that the effectiveness aspect achieved a good average score of 4.74, while the efficiency aspect obtained an average score of 4.55. Moreover, the user satisfaction aspect received a high average score of 4.92. Thus, the overall level of usability is rated as good, with an average usability score of 4.76. Accessibility elements such as multimedia, screen design, language, and object motion enhance the usability of serious games. With this testing, the prototype will be more usable. Therefore, this serious game meets the usability aspects of serious games for low vision children. In conclusion, this study shows that this serious game

is useful for children with low vision. The design of this game helps children play and learn more comfortably and improves their play experience. Future research could be done on designing an accessibility design for serious games for other disabilities, and new technology, such as artificial intelligence, could be implemented into the game design.

ACKNOWLEDGMENT

This research has received funding from the Ministry of Higher Education and the Faculty of Information Science and Technology, National University of Malaysia under FRGS Program (FRGS/1/2020/ICT03/UKM/02/4).

REFERENCES

- [1] S. Z. Mohid, "Model Reka Bentuk Ketercapaian Perisian Kursus Multimedia Untuk Kanak-Kanak Oku Pendengaran," Universiti Kebangsaan Malaysia, 2017.
- [2] R. R. Z. Ramli, N. Sahari, S. F. Mat Noor, M. M. Noor, N. A. Abd Majid, H. A. Dahlan, and A. N. Abd Wahab, "Assessing Usability of Learning Experience Prototype," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 9, pp. 20–36, 2022, doi: 10.3991/ijet.v17i09.29955.
- [3] S. R. Z. Abidin, S. F. M. Noor, and N. S. Ashaari, "Guidelines of brain-based learning through serious game for slow reader students," *Proc. 2017 6th Int. Conf. Electr. Eng. Informatics Sustain. Soc. Through Digit. Innov. ICEEI 2017*, vol. 2017-Novem, pp. 1–6, 2018, doi: 10.1109/ICEEI.2017.8312461.
- [4] Z. Cheng, F. Hao, and Z. Jianyou, "04. Research on Design of Serious Game Based on GIS Serious Game Engine Structure Based on GIS Serious game engine based on GIS is the combination of Access GIS to call required terrain data by OpenGIS," *Lect. Notes Comput. Sci.*, pp. 231–233, 2010.
- [5] World Health Organization, "Visual Impairment and Blindness," 2016.
- [6] N. I. Othman, N. A. M. Zin, and H. Mohamed, "Accessibility Requirements in Serious Games for Low Vision Children," *Proc. Int. Conf. Electr. Eng. Informatics*, vol. 2019-July, no. July, pp. 624–630, 2019, doi: 10.1109/ICEEI47359.2019.8988791.
- [7] M. M. Ali-Shahid and S. Sulaiman, "A case study on reliability and usability testing of a Web portal," *2015 9th Malaysian Softw. Eng. Conf. MySEC 2015*, pp. 31–36, 2016, doi: 10.1109/MySEC.2015.7475191.
- [8] N. A. A. Zaki, T. S. M. T. Wook, and K. Ahmad, "A usability testing of ASAH- for children with speech and language delay," *Proc. 2017 6th Int. Conf. Electr. Eng. Informatics Sustain. Soc. Through Digit. Innov. ICEEI 2017*, vol. 2017-Novem, pp. 1–6, 2018, doi: 10.1109/ICEEI.2017.8312392.
- [9] M. A. Hersh and B. Leporini, "An overview of accessibility and usability of educational games," *no. April 2016. 2013*. doi: 10.4018/978-1-4666-4422-9.ch005.
- [10] B. Lange, S. Flynn, and A. Rizzo, "Initial usability assessment of off-the-shelf video game consoles for clinical game-based motor rehabilitation," *Phys. Ther. Rev.*, vol. 14, no. 5, pp. 355–363, 2009, doi: 10.1179/108331909X12488667117258.
- [11] D. I. Rosli and M. Mohamad, "A usability testing on aa-ba-ta Arabic learning tool," *ICDLE 2010 - 2010 4th Int. Conf. Distance Learn. Educ. Proc.*, pp. 218–221, 2010, doi: 10.1109/ICDLE.2010.5606000.
- [12] A. Hussain, E. O. C. Mkpjojiogu, H. Abubakar, and H. M. Hassan, "The usability evaluation of Mudah.my on mobile device," *AIP Conf. Proc.*, vol. 1891, no. October, 2017, doi: 10.1063/1.5005391.
- [13] M. Verkuyll, L. Atack, P. Mastrilli, and D. Romaniuk, "Virtual gaming to develop students' pediatric nursing skills: A usability test," *Nurse Educ. Today*, vol. 46, pp. 81–85, 2016, doi: 10.1016/j.nedt.2016.08.024.
- [14] A. Syazwani, M. N. Siti Fadzilah, and M. Hazura, "SkillsMalaysia Journal Kebolegunaan Aplikasi M-Pembelajaran TVET," *Ski. J.*, vol. 4, no. 1, 2018, [Online]. Available: <http://www.ciaat.gov.my/journal%7C34-46>
- [15] N. B. N. Rozali and M. Y. B. Said, "Usability testing on government agencies web portal: A study on Ministry of Education Malaysia (MOE)

- web portal,” 2015 9th Malaysian Softw. Eng. Conf. MySEC 2015, pp. 37–42, 2016, doi: 10.1109/MySEC.2015.7475192.
- [16] M. S. Sulaiman and A. Azmi, “Interactive WebSIR Using Software Usability Measurement Inventory (SUMI),” *Adv. Electr. Comput. Sci. Electron. Eng. Comput. Sci.*, 2021.
- [17] T. T and A. T. Muharram, “The Analysis Knowledge Management System Of Electronic Government South Tangerang Based On Usability Evaluation Using SUMI (Software Usability Measurement Inventory),” *Data Sci. J. Comput. Appl. Informatics*, vol. 4, no. 1, pp. 47–58, 2020, doi: 10.32734/jocai.v4.i1-3203.
- [18] A. Donker and P. Reitsma, “Usability testing with young children,” *Proc. 2004 Conf. Interact. Des. Child. Build. a Community, IDC 2004*, pp. 43–48, 2004, doi: 10.1145/1017833.1017839.
- [19] M. Ismail, N. M. Diah, S. Ahmad, N. A. M. Kamal, and M. K. M. Dahari, “Measuring usability of educational computer games based on the user success rate,” *SHUSER 2011 - 2011 Int. Symp. Humanit. Sci. Eng. Res.*, pp. 56–60, 2011, doi: 10.1109/SHUSER.2011.6008500.
- [20] T. K. Chiew and S. S. Salim, “Webuse: Website usability evaluation tool,” *Malaysian J. Comput. Sci.*, vol. 16, no. 1, pp. 47–57, 2003.
- [21] T. Tullis and B. Albert, *Measuring the User Experience*. Elsevier, 2013, doi: 10.1016/C2011-0-00016-9.
- [22] K. K. Allah, N. A. Ismail, L. Hasan, and W. Y. Leng, “Usability Evaluation of Web Search User Interfaces from the Elderly Perspective,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 647–657, 2021, doi: 10.14569/IJACSA.2021.0121281.
- [23] P. Ghanouni, T. Jarus, J. G. Zwicker, and J. Lucyshyn, “An interactive serious game to Target perspective taking skills among children with ASD: A usability testing,” *Behav. Inf. Technol.*, vol. 40, no. 16, pp. 1716–1726, 2021, doi: 10.1080/0144929X.2020.1776770.
- [24] A. B. Najjar, A. Alhussayen, and R. Jafri, “Usability Engineering of a Tangible User Interface Application for Visually Impaired Children,” *Human-centric Comput. Inf. Sci.*, vol. 11, no. March, 2021, doi: 10.22967/HCS.2021.11.014.
- [25] M. Benaïda, “e-Government Usability Evaluation: A Comparison between Algeria and the UK,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, pp. 680–690, 2023, doi: 10.14569/IJACSA.2023.0140175.
- [26] G. P. Tolentino, C. Battaglini, A. C. V. Pereira, R. J. De Oliveria, and M. G. M. De Paula, “Usability of serious games for health,” *Proc. - 2011 3rd Int. Conf. Games Virtual Worlds Serious Appl. VS-Games 2011*, pp. 172–175, 2011, doi: 10.1109/VS-GAMES.2011.33.
- [27] Tan Wee Hoe, *Gamifikasi dalam Pendidikan*, 1st ed. Tanjung Malim: Universiti Pendidikan Sultan Idris, 2015.

Cybersecurity Advances in SCADA Systems

Machine Learning-based Insider Threat Detection and Future Directions

Bakil Al-Muntaser¹, Mohamad Afendee Mohamed², Ammar Yaseen Tuama³, Imran Ahmad Rana⁴

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia^{1,2}

College of Computer Science and Information Technology, University of Kirkuk, Iraq³

Superior University, Lahore, Pakistan⁴

Abstract—The management of critical infrastructure heavily relies on Supervisory Control and Data Acquisition [SCADA] systems, but as they become more connected, insider attacks become a greater concern. Insider threat detection systems [IDS] powered by machine learning have emerged as a potential answer to this problem. In order to identify and neutralize insider threats, this review paper examines the most recent developments in machine learning algorithms for insider IDS in SCADA security systems. A thorough analysis of research articles published in 2019 and later, focussed on variety of machine learning methods, have been adopted in this review study to better highlight difficulties and challenges being faced by professionals, and how the study will contribute to overcome them. The results show that, in addition to conventional methods, machine-learning based intrusion detection techniques offer important advantages in identifying complex and covert insider attacks. Finding pertinent insider threat data for model training and guaranteeing data privacy and security are still difficult to address. Ensemble techniques and hybrid strategies show potential for improving detection resiliency. In conclusion, machine learning-based insider IDS has the potential to protect critical infrastructures by strengthening SCADA systems against insider attacks. The similarities and differences between cyber physical systems and SCADA systems, emphasizing security challenges and the potential for mutual improvement were also reviewed in this study. In order to be as effective as possible, future research should concentrate on addressing issues with data collecting and privacy, investigating the latest developments in technology, and creating hybrid models. SCADA systems can accomplish proactive and effective defence against insider attacks by integrating machine learning advancements, maintaining their dependability and security in the face of emerging threats.

Keywords—Threat detection; SCADA security; machine learning-based intrusion detection; cyber-physical systems security; insider attack prevention

I. INTRODUCTION

SCADA systems are widely used in areas such as telecommunications, water management, electricity [1], [2]. These systems employ specialized control units like Master Terminal Units (MTUs) and Remote Terminal Units [RTUs] to automate and manage industrial operations [3]. With the advent of Industry 4.0, the integration of Internet of Things [IoT] devices into Cyber-Physical Systems (CPS) has given rise to Industrial IoT. This technology enables real-time monitoring of machine status, providing operators with immediate feedback and facilitating faster operations in industries that heavily rely on electrical machinery, particularly induction motors [1], [4].

Modernized SCADA systems have evolved into highly advanced and intricate technological systems. The adoption of open standard protocols has significantly enhanced their productivity and profitability. SCADA architecture offers numerous benefits, including improved data access, cost-effectiveness, flexibility, configurability, accessibility, and scalability [5], [7]. However, these advancements have also introduced new threats and vulnerabilities [6]. Similar concerns have been raised by other researchers [2], who emphasized the increasing cyber-attacks on SCADA systems due to their rapid evolution, automation, real-time operation, and decentralized, multi-component design. A few researchers [6],[4],[8] highlighted the complexity and severity of the situation, attributing it to the utilization of the Internet for communications within SCADA systems.

Scope of the Review: The study concentrated on investigation and analysis of recent developments in Insider Intrusion Detection Systems (IDS) in SCADA systems. A variety of machine learning methods, deep learning models, advanced algorithms, and other new techniques utilized for insider IDS in SCADA contexts were covered (Tables I to III). The studies have also examined the difficulties encountered when putting these ideas into practice and assess how well they work at identifying insider threats [9]. In this study the developments over the past five years have been discussed.

Research Problem: As the review study focuses on addressing key challenges by exploring the advancements and issues in insider threat detection systems (IDS) for SCADA environments. The research problem lies here is enhancing the detection and prevention of insider threats within SCADA systems using machine learning-based approaches. By examining the latest developments in this field, we aim to provide a comprehensive understanding of how machine learning techniques can be harnessed to fortify SCADA systems against insider attacks.

The significance of this study is twofold: Firstly, it offers insights into the application and efficiency of Insider IDS in SCADA systems, highlighting their capabilities to identify complex and covert insider attacks. Secondly, it delves into the limitations and obstacles faced while implementing these systems, guiding practitioners and researchers towards more effective and secure solutions. By bridging the gap between the evolving threat landscape and SCADA security, this review contributes to the advancement of cyber-physical system protection and ensures the continuous operation of critical infrastructures [8].

A. *Research Objectives: The Key Objectives of this Study are;*

- 1) to give a thorough review of insider IDS's application and efficiency in SCADA systems.
- 2) to recognize the main issues and new developments in insider threat detection and prevention in SCADA environments.
- 3) to assess the effectiveness and constraints of the current insider IDS systems and suggest additional areas for development.

B. *Research Questions: Based on above Objectives the Study Addressed following Questions;*

- 1) What are the prevalent methods and strategies employed in insider IDS for SCADA systems, and how have they evolved to address the growing challenges posed by insider attacks?
- 2) What are the key issues and challenges in insider threat detection and prevention in SCADA environments, and what are the recent advancements in machine learning-based intrusion detection techniques to overcome these challenges effectively?
- 3) How effective are the current insider IDS systems in SCADA environments, and what are their strengths and weaknesses? Additionally, what are the potential areas for further development to enhance the capabilities of insider IDS systems in SCADA?

This paper is structured as follows: In the Review of Literature section, we delve into the comprehensive body of existing research to establish the context and identify the gaps that motivate our study. The Review Methodology section outlines our approach to analyzing recent advancements and challenges in Insider IDS for SCADA systems. We have then presented our findings in the Discussion and Analysis section, where we explored the strengths, limitations, and potential solutions of different techniques. Then, at challenges and recommendations section, we addressed the identified gaps and provide insights into enhancing the effectiveness of Insider IDS. The Comparison of IDS SCADA and Cyber Physical section highlighted the distinctive features of both realms and their interaction [10]. Finally, we concluded by summarizing our key findings, followed by a discussion of limitations and future research directions.

By following this structure, we aim to offer a holistic understanding of the advancements and challenges in Insider IDS for SCADA systems, providing valuable insights to both researchers and practitioners in the field (Table IV).

II. REVIEW OF LITERATURE

A. *Machine Learning Techniques for Insider IDS in SCADA Systems*

SCADA systems are vital in maintaining and controlling different infrastructures, such as power grids, water treatment

facilities, and transportation networks. These systems come with higher risk of insider attacks as they become more networked and accessible [10], [11]. The security and dependability of SCADA systems are seriously threatened by insider assaults carried out by anyone with authorized access to the system.

Traditional security solutions frequently fall short in addressing these internal risks. Machine Learning [ML] based Insider Intrusion Detection Systems [IDS] have become a potential method for improving insider attack detection and prevention in SCADA systems. The machine learning is commonly known as a branch of artificial intelligence [12]. ML has the capacity to examine massive amounts of data, spot trends, and spot unusual behaviour that might be indicative of hostile intent.

For insider IDS in SCADA systems, machine learning techniques comprise on developing models from past data in order to understand typical system behaviour and spot abnormalities that can point to insider assaults. These methods benefit from flexibility and the capacity to learn from fresh facts, enabling them to advance alongside new dangers [13].

The use of machine learning in insider IDS enables the detection of intricate and covert attacks that rule-based or signature-based methods can miss. Machine learning algorithms may recognize small anomalies and identify aberrant actions that depart from established standards by learning patterns and behaviours from data.

It is not without issues, yet, to implement machine learning methods for insider IDS in SCADA systems [1], [14]. As recognized insider threat data is often difficult to come by due to the frequency of such instances, data collection, pre-processing, and classification can be challenging jobs. Furthermore, it is essential to protect the confidentiality and security of critical SCADA system data while developing machine learning models [14].

The goal of the current study is to examine the most recent developments in insider IDS techniques for SCADA systems. It has explored the many methodologies, approaches, and algorithms applied in this field (Tables I to II). The review has also gone through the advantages, difficulties, and potential future research paths of using insider IDS to improve the security and resilience of SCADA systems [15].

Insider IDS in SCADA systems may strengthen the ability to identify and respond against insider attacks by utilizing machine learning techniques. The subsections below will provide in-depth discussion on particular machine learning techniques and methodology used in insider IDS, as well as the difficulties and opportunities presented by their application in SCADA systems [16].

TABLE I. IDS SOLUTIONS COMPARISONS

IDS Solution	Algorithm Description	Scalability	Limitations	Advantages	Conditions for Algorithm Use
Snort	Rule-based Systems	Snort is suitable for small to medium-sized networks, easily scaled up by adding more hardware or utilizing distributed deployments [10].	Limited in detecting novel or zero-day attacks, requires regular updates to stay up-to-date with emerging threats [10].	Easy customization and quick deployment, widely used and well-established in various network environments.	Rule-based systems are the primary approach in Snort. Machine learning and statistical analysis are used to a limited extent for rule generation and identifying abnormal behavior. [10],[14]
Suricata	Rule-based Systems	Suricata is highly scalable and can handle large network traffic volumes, making it suitable for enterprise-level deployments.	Deployment complexity may require more expertise, continuous tuning for minimizing false positives [15].	Accurate detection of known attack patterns, data fusion capability enhances detection of complex attack scenarios.	Rule-based systems are the primary approach in Suricata. Data fusion is used to correlate data from multiple sources and enhance detection capabilities [15]. [16].
Bro/Zeek	Rule-based Systems, Statistical Analysis	Bro/Zeek is highly scalable and can handle large network traffic volumes, making it suitable for enterprise-level deployments.	Requires more computational resources due to the comprehensive analysis it performs, potential performance impact in high traffic scenarios.	Comprehensive network monitoring capabilities, rule-based approach with added statistical analysis for anomaly detection.	Rule-based systems are the primary approach in Bro/Zeek. Statistical analysis is used to identify anomalies and detect patterns within network traffic [19].20].
McAfee	Rule-based Systems, Machine Learning	McAfee IDS solutions are designed for scalability and can be deployed in various network environments, including small to large enterprise networks.	Complexity in managing machine learning models, potential false positives/negatives depending on training data quality.	Combination of rule-based systems with machine learning techniques for enhanced threat detection, suitable for different network environments [23].	Rule-based systems are the primary approach in McAfee IDS solutions. Machine learning is used to detect unknown or evolving threats in some versions of the solution [20],[25].
Cisco Firepower IPS	Rule-based, Statistical Analysis, Graph-based, Machine / Deep Learning, Data Fusion, NLP, Hybrid Approaches	Cisco Firepower IPS is designed to scale for large enterprise networks and can handle high network traffic volumes.	Complex deployment and management, potential resource-intensive processing for advanced techniques, dependence on quality training data.	Comprehensive set of detection techniques, ability to handle high network traffic volumes, advanced approaches like machine learning and graph-based analysis [14], [16].	Multiple algorithms are used based on the specific requirements and features. The solution employs rule-based systems, statistical analysis, machine learning, deep learning, graph-based approaches, and data fusion [23].

In order to solve the particular difficulties, problems, and traits of insider IDS [Intrusion Detection Systems] in SCADA systems, machine learning approaches have been developed and customized [17]. These modifications take into account the unique demands and limitations of SCADA systems, such as the necessity for rapid and precise anomaly detection and real-time monitoring and large-scale data processing. We have gone through some of the most significant modifications made to machine learning methods for insider IDS in SCADA systems in this part.

- **Feature Engineering:** In SCADA systems, feature engineering is key to machine learning for insider IDS. Domain-specific features must be identified and designed due to the nature of SCADA data, which consists of time-series measurements, sensor readings, and control orders [18]. These characteristics record crucial facets of system activity and give machine learning models useful input. Statistical measurements, signal processing methods, and frequency domain

analysis are a few examples of features frequently found in SCADA systems.

- **Imbalanced Data:** Compared to typical system behaviour, insider attacks are frequently infrequent occurrences. As a result, datasets become unbalanced, with a vastly greater number of regular instances than attacks [1]. Machine learning algorithms may have trouble effectively identifying the minority class of attacks. The performance of machine learning algorithms can be strengthened by rebalancing the dataset using a variety of techniques, such as oversampling, under sampling, or the use of ensemble methods like SMOTE (Synthetic Minority Over-Sampling Technique).
- **Real-time Processing:** SCADA systems work in real-time settings where it's crucial to promptly detect and respond to insider threats. For SCADA systems to handle the high-speed data streams they produce, machine learning algorithms must be modified. Models can update and react in real-time thanks to methods like

online learning or streaming algorithms, ensuring the prompt detection of insider threats [19].

- **Model Interpretability:** In the context of SCADA systems, machine learning model interpretability is crucial. The reasoning behind the judgments made by the models must be understood by system administrators and security staff. Rule extraction, feature importance analysis, or the application of explainable AI [XAI] methodologies are a few techniques that can give insights into model behaviour and improve user confidence in and comprehension of the applied machine learning models [9].
- **Resource Constraints:** SCADA systems frequently use constrained computational resources, necessitating the development of effective and portable machine learning models. Reduce the computing requirements of the models without sacrificing performance by using methods like model compression, model pruning, or the usage of simplified architectures [10].
- **Transfer Learning and Domain Adaptation:** Transfer Learning and Domain Adaptation methods are useful because classified insider threat data is hard to get in SCADA systems [10]. Retrained models from related domains can be used as a starting point or to bootstrap the training process, which may then be fine-tuned using the specific data from the SCADA system. By using this strategy, the machine learning models perform better and the problem of data scarcity is lessened.

The improvements and adjustments of machine learning approaches for insider threats in SCADA systems take into account the particular difficulties and traits of these systems. They make it possible to create intrusion detection systems that are effective and efficient, capable of handling unbalanced data, providing interpretability, and operating within the resource limitations of SCADA environments [12]. In order to strengthen the security of SCADA systems against insider assaults, current research in this area intends to further improve and refine these adaptations along with investigating fresh approaches.

Insider IDS [Intrusion Detection Systems] for SCADA [Supervisory Control and Data Acquisition] systems provide compelling potential for fresh methodologies and applications. These methods improve insider threat detection and prevention by utilizing the capabilities of machine learning algorithms. We'll talk about a few cutting-edge methods and uses of machine learning in this part as they relate to insider IDS for SCADA systems [5], [7].

- **Behaviour Analysis:** Using machine learning algorithms for behavioural analysis in SCADA systems is one cutting-edge strategy. Machine learning models can identify variations that can be signs of insider assaults by learning the typical patterns of user behaviour and system operations [10]. To respond to changing system behaviour, these models can be continuously updated and trained using past data.

- **Anomaly Detection:** Machine learning approaches are excellent at recognizing anomalies, which is essential for identifying insider threats in SCADA systems [9]. Models are able to spot alterations from the norm that could be indicative of malicious activity by learning from the typical system behaviour. Auto encoders, Gaussian mixture models, or one-class SVMs are examples of unsupervised learning techniques that can be used to quickly identify anomalies and flag questionable behaviour [10].
- **Ensemble methods:** To increase detection accuracy and robustness, ensemble approaches mix various machine learning models. Ensembles of models can be built using methods like bagging, boosting, or stacking that were trained on various subsets of the data or with various techniques. This ensemble-based strategy reduces false positives and improves insider IDS's overall performance in SCADA systems [12], [16].

B. Deep Learning Models

For insider IDS in SCADA systems, deep learning models like deep neural networks, have demonstrated potential in a number of fields. These models are capable of discovering complex features and patterns from unprocessed data, which makes it possible to identify sophisticated insider attacks [13]. For evaluating network traffic and time-series data, Convolutional Neural Networks [CNNs] and Recurrent Neural Networks [RNNs] are frequently used in deep learning-based IDS.

- **Graph-Based Approaches:** SCADA systems frequently have a complicated network structure, where components and dependencies can be represented as graphs. The relationships between system items can be captured and potential insider threats can be identified using graph-based machine learning techniques (Tables II and III). In SCADA systems, suspicious patterns, attack pathways, and alert priority can all be found using methods like graph neural networks and graph clustering techniques [17].
- **Hybrid Approaches:** To take use of their complimentary qualities, hybrid approaches incorporate various machine learning techniques such as rule-based systems, statistical analysis, and machine learning algorithms. These methods make it possible to combine several detection techniques, and they improve insider IDS in SCADA systems' precision and robustness [12].
- **Natural Language Processing [NLP]:** Textual information from logs, configuration files, and system messages can offer crucial information for spotting insider threats in SCADA systems according to Natural Language Processing [NLP]. The NLP techniques can be used to process and analyse this textual data, extract pertinent data, and spot potential attack indications.

A thorough insider IDS architecture can benefit from the use of machine learning models that can be trained to categorize and analyze text input [18].

These revolutionary methods and insider IDS for SCADA systems uses of machine learning techniques offer enormous potential to improve security and safeguard crucial infrastructures. Continuous research and development in these field aims to improve the effectiveness of these methods, handle the changing problems brought on by insider threats in SCADA systems, and further these approaches [6].

In addition to the machine learning techniques, deep learning models, and advanced algorithms mentioned earlier, there are other tools and approaches that can be explored in the context of insider IDS in SCADA systems (Table IV). Here are a few examples:

- **Statistical Approaches:** Data from SCADA systems can be examined for patterns, trends, and abnormalities using statistical approaches. Techniques like time series analysis, statistical process control, and multivariate analysis can shed light on unusual behaviour or departures from the way a system is supposed to work [7].
- **Rule-based Systems:** To identify potential insider threats, rule-based systems use a set of established rules or criteria. These guidelines may be drawn from professional judgment or widely accepted commercial norms. To improve detection accuracy, rule-based systems are frequently employed in conjunction with other methods [16].
- **Data Fusion:** To increase the precision of insider threat detection, data fusion merges data from many sources, including sensor data, network logs, and user behaviour. Complex patterns and correlations that might not be obvious when evaluating individual data sources might be found by integrating various data streams [20].
- **Evolutionary Algorithms:** To maximize a solution, evolutionary algorithms imitate the processes of natural selection and genetic evolution. They can be used to improve insider IDS parameters, feature choices, or model architectures in SCADA systems. Examples of evolutionary algorithms include differential evolution, genetic algorithms, and particle swarm optimization [4].
- **Reinforcement learning** includes teaching an agent how to make decisions sequentially in a setting to maximize a reward signal. This method can be used to develop IDS systems that are self-adaptive and self-learn in order to dynamically respond to insider threats in SCADA systems [19].
- **Graphical probabilistic models** called Bayesian networks are used to describe ambiguous relationships between variables. They can be used to simulate causal chains and interdependencies within a SCADA system, making it easier to spot irregularities and possible insider threats.
- **Fuzzy Logic** is a mathematical framework for handling deliberation and decision-making in the face of uncertainty. In insider IDS for SCADA systems, it can be used to describe imprecise or uncertain knowledge,

enabling more adaptable and reliable detection procedures [13], [19].

This broadens the selection of tools available for identifying insider threats in SCADA systems by offering alternatives to conventional machine learning and deep learning techniques. When choosing and implementing these approaches, it's crucial to thoroughly evaluate their applicability for the specific SCADA environment and take into account their advantages, disadvantages and performance traits.

C. Cyber Physical Systems

Cyber-Physical Systems (CPS) are networked systems that combine computing and communication skills with physical aspects [10]. CPS denotes to the integration of SCADA systems with the latest information and communication technologies in the context of this review study. CPS is essential for controlling critical infrastructure, but it also creates new security risks, especially with regard to insider attacks [21],[22]. To increase the security of CPS, the study intends to investigate the most recent advancements and breakthroughs in Insider Intrusion Detection Systems (IDS) based on machine learning. The study seeks to pinpoint major issues and offer suggestions for proactive security against insider assaults in CPS contexts by examining the use and efficacy of various machine learning approaches (Table III).

III. IDS SOLUTIONS

Machine learning is a subfield of artificial intelligence (AI) which entails training computers to learn from data and make judgments or predictions without being explicitly programmed. ML approaches can be applied to Intrusion Detection Systems (IDS) to improve the solutions' capacity for identifying potential hacking attempts or anomalies [5].

The major components of the IDS solutions like Snort, Suricata, Bro/Zeek, McAfee IPS, and Cisco Firepower IPS—use established rules or patterns to identify known threats. These IDS systems can, however, be enhanced using ML algorithms to make them more intelligent and flexible [15],[23].

An outline of how ML relates to IDS solutions is provided below: Data is necessary for ML algorithms to learn from and generate predictions with. In the case of IDS, ML models can be trained using historical network traffic data [24].

Learning Patterns: Machine learning algorithms examine the training data and discover patterns or traits that distinguish between legitimate and harmful activity. For instance, they can spot specific network traffic patterns linked to particular kinds of attacks [13].

Making Predictions: Following training, ML models can be used to generate predictions about incoming network traffic that hasn't yet been seen. The models look for any signs of an ongoing assault or unusual activity by comparing the observed traffic patterns with what they have learnt from the training data.

Adaptability and Anomaly Detection: ML algorithms are also capable of identifying anomalies, which are atypical

patterns or behaviours that don't correspond to well-known assault patterns. Due to their versatility, ML-based IDS solutions can recognize new or undiscovered assaults [12], [13].

Continuous Improvement: As they are exposed to more data over time, ML models can continuously train and get better. They can refresh their knowledge to recognize new risks and respond to changing attack methodologies [12].

In summary, ML algorithms are employed in IDS solutions to learn from historical network traffic data, discover patterns linked to legitimate and nefarious conduct, and generate

forecasts about impending attacks or anomalies. This makes IDS solutions more proficient in identifying and stopping network intrusions.

It's important to note that the conditions for algorithm use as displayed in Table II are general guidelines which may vary based on the specific configurations, versions, and deployment environments of each IDS solution. The performance metrics for each technique, such as detection accuracy, false positive rate, and response time, are compared thoroughly in the Table II. These were selected based on studies [26], [27].

TABLE II. PERFORMANCE METRICS REVIEW FOR INSIDER IDS SCADA SYSTEMS TECHNIQUES

Technique	Detection Accuracy	False Positive Rate	Response Time	Applicability / Usage Condition	Limitation	Advantages	Other Information
Behavioural Analysis	High	Low	Fast	Effective for identifying insider threats	May not capture all insider attack patterns	Captures unusual patterns of behavior that may indicate insider attacks [1],[3]	Measures the accuracy of detecting insider threats based on behavior
Anomaly Detection	High	Low	Fast	Effective for identifying unknown threats	May result in false positives for certain data distributions	Can detect previously unseen insider attack patterns	Measures the accuracy of detecting anomalies in network traffic [3]
Ensemble Methods	High	Low	Fast	Effective for improving overall accuracy	Complexity may lead to higher resource requirements [9]	Combines multiple models to reduce false positives and increase accuracy [14]	Measures the overall detection accuracy of the ensemble
Deep Learning Models	High	Low	Fast	Effective for complex pattern recognition	Requires large amounts of labelled data	Can identify intricate patterns and detect novel insider threats	Measures the accuracy of detecting threats based on deep learning models
Graph-Based Approaches	High	Low	Fast	Effective for capturing network relationships	May face challenges in large-scale networks	Can detect insider threats by analyzing complex relationships in SCADA	Measures the accuracy of detecting threats based on graph analysis [15]
Hybrid Approaches	High	Low	Fast	Effective for improving overall accuracy	Requires careful integration of different techniques	Combines multiple methods to enhance detection capabilities	Measures the overall detection accuracy [14],[15]
Natural Language Processing [NLP]	High	Low	Fast	Effective for analysing textual data	Requires pre-processing of unstructured data	Can identify indicators of potential insider attacks from text data [21]	Measures the accuracy of detecting insider threats based on NLP

IV. REVIEW METHODOLOGY

A structured approach has been used in the review study to collect, evaluate, and synthesize relevant research on insider IDS in SCADA systems. The methodology is comprised on following steps:

Literature Search: To find studies, conferences, and journal publications from the previous five years [2019 to the present] that are relevant to the objectives of the research, a thorough search was carried out in research databases including IEEE Xplore, WOS, ScienceDirect, and Google Scholar.

The selection of relevant research is based on previously established inclusion and exclusion criteria. This study has taken into account works that concentrate on machine learning methods, deep learning models, advanced algorithms, and new insider IDS strategies in SCADA systems. Papers that don't fit the criteria for scope or quality were not being considered.

Data Extraction: Important details from the chosen articles, including the title, authors, publication year, methodology, algorithms employed, performance measures, and conclusions were taken out. In order to facilitate comparison and analysis, such information was displayed in a tabular manner.

The search was conducted in June 2023, and the years 2019 through 2023 were taken into consideration. 183 papers were obtained as a result, including: 59 papers from Science Direct, 77 papers from WOS, Google scholar, and 47 papers from IEEE Xplore. The final collection contained 94 papers after manual inspection and the elimination of the replicated publications.

Additionally, we chose potential articles based on the titles and abstracts, paying particular attention to those that offered novel suggestions for NIDS-specific to SCADA [5],[14]. There were 51 papers in the remaining collection.

After reviewing the complete candidate papers, the goal was to collect a set of original and similar solutions. As a

result, we disregarded publications that were similar in content but had different authors or described the outcomes of the same initiatives, as well as papers that lacked IDS evaluation findings. 27 papers were selected in the end.

Data Analysis: To find patterns, trends, and insights regarding the most recent developments in insider IDS for SCADA systems, the retrieved data was evaluated and synthesized. A comparative analysis was done to assess how

well various strategies and algorithms perform [Reference Table IV].

Development of a Conceptual Framework: Based on the research and analysis, conceptual framework was constructed to classify and comprehend the numerous machine learning methods, deep learning models, and advanced algorithms utilized in insider IDS for SCADA systems (Table III).

TABLE III. CPS AND SCADA SECURITY LANDSCAPE AND LEARNINGS REVIEW

Comparison Aspect	Cyber-Physical Systems [CPS]	SCADA Systems	Security Landscape Comparison	Advantages for Learning
Integration of Technologies	CPS integrates cyber and physical components to create interconnected systems [6].	SCADA focuses on integrating sensors, actuators, and control systems in industrial environments.	Both CPS and SCADA require secure integration of diverse technologies to prevent cyber-physical attacks and ensure data integrity.	CPS can learn from SCADA's focus on industrial protocols and network segmentation to enhance security. SCADA can learn from CPS's advanced encryption and authentication techniques for securing data flow in integrated systems [17].
Data Collection and Analysis	CPS relies on data from sensors and other sources for real-time monitoring and control.	SCADA systems gather data from sensors and devices to monitor industrial processes.	Both CPS and SCADA must secure data collection and analysis to prevent unauthorized access or tampering.[20]	CPS can learn from SCADA's data filtering techniques for efficiently handling large data streams. SCADA can learn from CPS's data analytics capabilities to improve predictive maintenance and anomaly detection [21].
Real-time Monitoring and Control	CPS provides real-time feedback and control in various domains.	SCADA allows operators to monitor and control industrial processes in real-time.	Both CPS and SCADA face real-time security challenges, requiring robust authentication and authorization mechanisms.	CPS can learn from SCADA's focus on redundancy and fail-safe mechanisms for continuous real-time control. SCADA can learn from CPS's distributed control architecture for improved system resiliency [22].
Connectivity and Communication	CPS uses communication networks to facilitate data exchange between cyber and physical components.	SCADA relies on communication networks for data transmission between central control and field devices.	Both CPS and SCADA must ensure secure communication channels to prevent unauthorized access and data interception.	CPS can learn from SCADA's strict access control policies and encryption techniques for secure communication. SCADA can learn from CPS's adaptive communication protocols for handling dynamic network conditions [22].
Security Challenges	CPS and SCADA face security challenges due to their interconnected nature.	Ensuring data and communication security is crucial to prevent cyber-attacks and disruptions.	Both CPS and SCADA must address security challenges related to insider threats, remote access, and supply chain vulnerabilities [21].	CPS can learn from SCADA's robust anomaly detection mechanisms for detecting suspicious activities. SCADA can learn from CPS's threat intelligence integration for proactive identification of potential cyber threats.
Industrial Applications	CPS finds applications in manufacturing, energy, healthcare, transportation, etc.	SCADA is commonly used in industrial sectors for process control and automation.	Both CPS and SCADA play critical roles in enhancing efficiency, automation, and optimization of processes in their respective domains [22].	CPS can learn from SCADA's domain-specific protocols and standards for seamless integration into industrial applications. SCADA can learn from CPS's adaptability to diverse industrial settings for improved flexibility and scalability.

TABLE IV. IDS ALGORITHMS REVIEW

Algorithm	Usage Conditions	Core Strengths	Core Weaknesses	Opportunities	Core Threats	Supporting Technologies & Industries
Ensemble Methods	Diverse dataset, multiple models	Improved prediction accuracy, model robustness	Increased complexity and computational resources [4]	Ensemble learning techniques, model combination	Overfitting, model selection, ensemble diversity	Various machine learning frameworks, successful in various industries such as finance, healthcare, and retail
Evolutionary Algorithms	Complex optimization problems	Effective for global optimization, handle constraints	Computationally expensive, slow convergence	Optimization of insider IDS parameters and features [10], [11]	Premature convergence, parameter tuning	Genetic algorithms, particle swarm optimization [7]
Hybrid Models	Diverse techniques, flexible	Combines strengths of different approaches	Complexity in model integration, interpretability trade-off [10],[13]	Improved detection accuracy, adaptable systems	Increased complexity, model integration challenges [20]	Integration of machine learning and rule-based systems
Anomaly Detection	Unusual patterns, outlier detection	Identifies unknown and rare insider threats	Difficulty in defining normal behaviour, high false positives	Uncovering novel insider threats, pattern recognition	Sensitivity to data quality, evolving threats	Statistical analysis, unsupervised learning, successful in various industries such as cybersecurity and fraud detection [12], [13]
Graph-based Methods	Network or relationship data	Captures complex relationships, detects structural anomalies	High computational cost for large graphs, graph construction	Identifying suspicious connections, network analysis	Scalability, graph sparsity, noise in data	Network analysis tools, successful in social networks, cybersecurity, and transportation systems [7],[20]
Reinforcement Learning	Sequential decision-making tasks	Adapts to dynamic environments, learns from interactions	High sample complexity, sensitivity to reward design	Adaptive and self-learning IDS systems	Exploration-exploitation trade-off, reward design	Q-learning, deep reinforcement learning frameworks
Bayesian Networks	Uncertain relationships, probabilistic reasoning	Captures causal dependencies, handles uncertainty	Requires prior knowledge, limited scalability	Modelling complex relationships, uncertainty handling	Learning structure from data, complexity in learning	Probabilistic programming, successful in medical diagnosis, risk assessment, and fault diagnosis [28], [29]
Fuzzy Logic	Handling imprecise knowledge, reasoning under uncertainty	Captures vague and uncertain information	Interpretability, intuitive reasoning	Modelling linguistic variables, fuzzy rule-based systems	Knowledge representation, fuzzy rule tuning	Fuzzy logic controllers, successful in industrial automation, decision support systems, and robotics [23]

V. CHALLENGES AND RECOMMENDATIONS

A. Challenges

Supervisory Control and Data Acquisition [SCADA] systems must be protected from internal threats by insider intrusion detection systems [IDS]. Insider attacks are more likely as SCADA systems grow more digitalized and networked, therefore strong and flexible security measures are required. This article examines the most recent difficulties encountered by Insider IDS in SCADA systems and makes suggestions to improve their efficiency and resiliency.

Data overload: Sensors, control systems, and communication networks all contribute to the massive volumes of data that SCADA systems produce. Insider IDS faces substantial challenges in processing and evaluating this data in real-time because doing so calls for strong computational capabilities and effective data handling techniques [1].

Anomaly Detection in Complex Environments: In the extremely dynamic and complex SCADA environment, insider IDS must be able to differentiate between legal deviations and probable insider assaults. Finding the right balance between detecting real anomalies and setting out false alarms is difficult, and SCADA operations vary widely [30].

Zero-Day assaults: Conventional signature-based detection techniques may have difficulty spotting new and unidentified assaults, such as zero-day threats. Insiders can take advantage of previously unknown flaws, making more sophisticated detection methods that go beyond specified rules necessary [8],[9].

Unbalanced Data Distribution: In SCADA systems, legitimate activity outweighs criminal activity by a large margin. The Insider IDS algorithms may be biased toward usual patterns as a result of this imbalance, which could affect how accurately they detect threats [1].

Limited Access to Training Data: Due to the sensitive nature of SCADA systems, obtaining labeled training data for Insider IDS can be difficult. Accurate machine learning models could be difficult to construct because of data access limits and privacy issues.

Adaptability to Evolving Attacks: Insider IDS must change in order to detect new and sophisticated threats as insider attack strategies change. To keep the system resilient to new attack vectors, regular upgrades and ongoing learning are essential [18].

SCADA systems need real-time monitoring and reaction capabilities in order to quickly minimize insider threats. The

integrity and safety of critical infrastructure can be severely compromised by any lag in detection and reaction.

B. Recommendations to Address Challenges:

Machine Learning-Based Detection: To improve Insider IDS capabilities, use machine learning techniques like deep learning models and anomaly detection algorithms. Even in complicated SCADA setups, these techniques may learn from prior data and spot patterns that can point to insider assaults [30].

Effective data preparation and feature engineering are essential to overcoming the problems brought on by data overload and unbalanced distributions. Insider IDS's accuracy can be improved by selectively choosing and extracting the most pertinent features [7].

Implement mechanisms for machine learning models that allow for continuous training and updating. The system's ability to respond to changing insider threats is enhanced by routinely providing it with fresh data.

Encourage cooperation between SCADA operators, vendors, and cybersecurity professionals so they can share threat intelligence and experiences. Collaboration can result in the discovery of fresh attack pathways and the creation of more potent detection methods.

Adopt hybrid IDS strategies that incorporate rule-based systems, statistical analysis, and machine learning techniques. Utilizing the advantages of several detection techniques can improve detection precision and lower false positives [1],[26].

User Monitoring and Behavioral Profiling: Use behavioral profiling of users and administrators to find alterations in normal patterns of activity. User monitoring can offer insightful information about shady behavior and possible insider threats [17].

Real-Time Incident Response: Create real-time response plans to quickly stop insider attacks. SCADA systems can avoid future harm with automated reactions like isolating hacked devices or halting unauthorized activity.

VI. COMPARISON OF IDS SCADA AND CYBER PHYSICAL SYSTEMS (TABLE III)

Cyber-Physical Systems [CPS] and SCADA systems are similar and dissimilar in many ways such as the integration of technologies, data collecting and analysis, real-time monitoring and control, connectivity and communication, security issues, and industrial applications. Due to their interconnected nature, CPS and SCADA both need to handle security issues such supply chain vulnerabilities and insider threats [22]. Additionally, for these systems to be protected from cyber-physical threats, unauthorized access, safe integration of various technologies and communication networks is essential.

The strengths of CPS and SCADA can be used to improve each other's security and effectiveness. The emphasis on industrial protocols and network segmentation in SCADA can help CPS, and SCADA can gain knowledge from CPS's cutting-edge encryption and authentication methods for securing data flow in integrated systems. Additionally, SCADA can benefit from CPS's distributed control architecture

and adaptive communication protocols for increased resiliency and flexibility in a variety of industrial settings, while CPS can learn from SCADA's data filtering strategies and emphasis on redundancy for improved efficiency and continuous real-time control [20],[21].

CPS and SCADA can improve their capabilities and security by sharing best practices and leveraging one another's strengths, so enhancing the general effectiveness, safety, and dependability of industrial processes and cyber-physical systems [6], [17], [20].

VII. DISCUSSION AND ANALYSIS

The study has conducted a comprehensive analysis of developments, obstacles, and efficiency to strengthen SCADA security against insider attacks. Examining prevalent techniques like behavioral analysis and anomaly detection, the research identifies their advantages over others for identifying insider threats (Table II). It also reveals barriers to insider IDS integration, such as the difficulty of interpreting models and the lack of available data. Encouraging advancements like explainable AI and federated learning are discussed that may open up new possibilities for cooperative threat detection [30].

The part on research objectives offers a thorough analysis of the use of insider IDS, new advancements, and efficacy evaluation. For researchers and professionals looking to strengthen vital infrastructures, it delivers insightful information.

The primary goal of this study was to offer a thorough analysis of the use and efficacy of insider IDS in SCADA systems. The study discovered that machine learning-based IDS systems have developed significantly to meet the expanding problems encountered by insider threats through a thorough analysis of the most recent developments [11], [12], [14]. These machine learning methods are good at interpreting massive volumes of data, recognizing patterns, and identifying unusual behavior that can point to malicious intent.

The review covered a wide range of machine learning techniques, such as natural language processing [NLP], graph-based approaches, ensemble methods, deep learning models, anomaly detection, behavioral analysis, and ensemble methods. The benefits and drawbacks of each method for identifying insider attacks on SCADA systems were examined (Tables II and III).

Important findings from studies revealed that behavioral analysis, applying machine learning algorithms, successfully learned typical patterns of user behavior and system operations, detecting variations suggestive of insider threats [10], [12]. Another machine learning-driven strategy called anomaly detection, which takes its nods from typical system activity and looks for abnormalities from established patterns, proved essential in identifying complex attacks.

The second objective of the study was to determine whether current insider threat detection and mitigation [IDS] techniques were appropriate for SCADA systems. The study indicated that insider attacks present certain complications that conventional security measures often are unable to meet. Machine learning-based systems demonstrated their capability

for learning from new data and adaptability, continuously improving alongside new threats [5], [11], [29] (Tables I and IV).

The ability to integrate different machine learning models using ensemble approaches like bagging, boosting, and stacking has been shown to improve detection accuracy and boost insider IDS in SCADA systems [2],[6],[8]. The accuracy and robustness of IDS solutions were further enhanced by the hybrid techniques that included various detection techniques.

The third objective of the review study was to find the key challenges and constraints encountered while using insider IDS solutions in SCADA systems. Due to the frequency of such incidents, the study noted the challenge in obtaining recognized insider threat data, making data collection, pre-processing, and classification difficult jobs. The review study findings reveal that it may be addressed by integrating Statistical Analysis techniques with rule based methods in IDS solutions (Tables I and III).

It also emphasized that it's very crucial to protect the security and confidentiality of crucial SCADA system data when creating machine learning models. In order to increase the transparency and interpretability of machine learning-driven IDS solutions and make sure that professionals can understand and trust the judgments made by these models, the review emphasized the necessity to integrate explainable AI. For greater transparency and confidence in the detection process, explainable AI needs to be integrated with IDS solutions in SCADA systems. Potential approaches that can accomplish this goal include ensemble methods, graph-based approaches, and rule-based systems. As they provide concrete concepts for decision-making, rule-based systems are simple to understand [27]. By revealing specific model contributions, ensemble approaches, and explicable AI techniques increase transparency. Understanding complicated relationships and attack pathways is made possible by graph-based techniques and explainable AI [30], [31]. By utilizing these methods, IDS systems in SCADA can deliver precise, comprehensible data, improving security analysts' capacity to recognize and efficiently address insider threats.

VIII. CONCLUSION, LIMITATIONS AND FUTURE DIRECTIONS

In conclusion, the discussion and analysis of this study offered valuable insight into the most recent developments, advantages, and difficulties of insider IDS in SCADA systems. The outcomes highlighted the significance of utilizing machine learning techniques as well as their adaptability and potential to change SCADA security. The review not only advanced academic knowledge in this field but also provided practitioners seeking to strengthen the security of vital infrastructures with practical recommendations (Tables II to IV).

IX. LIMITATIONS

Limited Scope: The study excludes other alternative strategies and technologies that can improve insider threat detection in favor of Insider Intrusion Detection Systems [IDS] for SCADA systems.

1) **Data availability:** An important limitation was the lack of pertinent insider threat data available for model training and evaluation. The findings' applicability to other situations may be restricted by the dearth of data from actual insider attacks.

2) **Time restrictions:** The study only included research papers published in 2019 and later, which may have left out some pertinent studies conducted earlier.

3) **Generalization:** Due to differences in system architectures, data formats, and threat settings, the conclusions and suggestions may not be universally applicable to all SCADA environments.

X. FUTURE DIRECTIONS

Enhanced Data acquiring: To improve the training and assessment of machine learning models, future research might concentrate on acquiring more thorough and varied insider threat data.

Explainable AI Integration: To improve the understand ability and transparency of machine learning-driven IDS solutions, more research can study and apply cutting-edge explainable AI methodologies.

Implementation in the real world: It would be helpful to conduct field tests and case studies to evaluate the usefulness and efficacy of machine learning-based IDS in actual SCADA environments.

Hybrid ways: Researching and creating hybrid ways that combine the benefits of various techniques, such as ensemble methods and rule-based systems, may result in the detection of insider threats being more reliable and accurate.

Exploring federated learning strategies, which enable cooperative model training across several SCADA systems without exchanging sensitive data, may be a promising step toward resolving data privacy issues.

Resistance to Adversarial Attacks: It would be beneficial to conduct research to make machine learning models more resistant to adversarial attacks, which might be used by knowledgeable insiders.

Collaboration between industries: Working together with SCADA system manufacturers and industrial groups could make it easier for machine learning-based IDS solutions to be adopted in practical environments.

The field of insider IDS for SCADA systems can go further by resolving these constraints and pursuing the indicated future directions, protecting critical infrastructures from the growing threat of insider threats.

XI. CONFLICT OF INTEREST STATEMENT

The authors affirm that they have no known conflicts of interest that would have appeared to have an impact on the research presented in this study.

REFERENCES

- [1] Balla, A., Habaebi, M. H., Elsheikh, E. A., Islam, M. R., & Suliman, F. M. [2023]. The Effect of Dataset Imbalance on the Performance of SCADA Intrusion Detection Systems. *Sensors*, 23[2], 758.

- [2] Salahudin, F., & Setiyono, B. [2019, September]. Design of Remote Terminal Unit [RTU] Panel Supply Monitoring Based on IOT Case Study at PLN. In 2019 6th International Conference on Information Technology, Computer and Electrical Engineering [ICITACEE] [pp. 1-6]. IEEE.
- [3] Elhady, A. M., El-Bakry, H. M., & Abou Elfetouh, A. [2019]. Comprehensive risk identification model for SCADA systems. *Security and Communication Networks*, 2019.
- [4] Huang, J. C., Zeng, G. Q., Geng, G. G., Weng, J., Lu, K. D., & Zhang, Y. [2023]. Differential evolution-based convolutional neural networks: An automatic architecture design method for intrusion detection in industrial control systems. *Computers & Security*, 132, 103310.
- [5] Öztürk, T., Turgut, Z., Akgün, G., & Köse, C. [2022]. Machine learning-based intrusion detection for SCADA systems in healthcare. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 11[1], 47.
- [6] Alanazi, M., Mahmood, A., & Chowdhury, M. J. M. [2022]. SCADA vulnerabilities and attacks: A review of the state-of-the-art and open issues. *Computers & Security*, 103028.
- [7] Qian, J., Du, X., Chen, B., Qu, B., Zeng, K., & Liu, J. [2020]. Cyber-physical integrated intrusion detection scheme in SCADA system of process manufacturing industry. *IEEE Access*, 8, 147471-147481.
- [8] Livinus Obiora Nweke, "A Survey of Specification-based Intrusion Detection Techniques for Cyber-Physical Systems" *International Journal of Advanced Computer Science and Applications*[IJACSA], 12[5], 2021. <http://dx.doi.org/10.14569/IJACSA.2021.0120506>
- [9] Saheed, Y. K., Abdulganiyu, O. H., & Tchakoucht, T. A. [2023]. A Novel Hybrid Ensemble Learning for Anomaly Detection in industrial sensor networks and SCADA systems for smart city infrastructures. *Journal of King Saud University-Computer and Information Sciences*, 35[5], 101532.
- [10] Alem, S., Espes, D., Nana, L., Martin, E., & De Lamotte, F. [2023]. A novel bi-anomaly-based intrusion detection system approach for industry 4.0. *Future Generation Computer Systems*.
- [11] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. [2023]. Machine Learning-based Intrusion Detection for Smart Grid Computing: A Survey. *ACM Transactions on Cyber-Physical Systems*, 7[2], 1-31.
- [12] Mohammad Azmi Ridwan, Nurul Asyikin Mohamed Radzi, Kaiyisah Hanis Mohd Azmi, Fairuz Abdullah and Wan Siti Halimatul Munirah Wan Ahmad, "A New Machine Learning-based Hybrid Intrusion Detection System and Intelligent Routing Algorithm for MPLS Network" *International Journal of Advanced Computer Science and Applications*[IJACSA], 14[4], 2023. <http://dx.doi.org/10.14569/IJACSA.2023.0140412>
- [13] Yang, H., Cheng, L., & Chuah, M. C. [2019, June]. Deep-learning-based network intrusion detection for SCADA systems. In 2019 IEEE Conference on Communications and Network Security [CNS] [pp. 1-7]. IEEE.
- [14] Sivakumar, S., Raffik, R., Kumar, K. K., & Hazela, B. [2023, January]. Scada energy management system under the distributed decimal of service attack using verification techniques by IIoT. In 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering [ICECONF] [pp. 1-4]. IEEE.
- [15] Khan, I. A., Pi, D., Khan, Z. U., Hussain, Y., & Nawaz, A. [2019]. HML-IDS: A hybrid-multilevel anomaly prediction approach for intrusion detection in SCADA systems. *IEEE Access*, 7, 89507-89521.
- [16] Bugshan, N., Khalil, I., Kalapaaking, A. P., & Atiquzzaman, M. [2023]. Intrusion Detection-Based Ensemble Learning and Microservices for Zero Touch Networks. *IEEE Communications Magazine*, 61[6], 86-92.
- [17] Asiri, M., Saxena, N., Gjomemo, R., & Burnap, P. [2023]. Understanding indicators of compromise against cyber-attacks in industrial control systems: a security perspective. *ACM transactions on cyber-physical systems*, 7[2], 1-33.
- [18] Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. [2021]. A review of research works on supervised learning algorithms for SCADA intrusion detection and classification. *Sustainability*, 13[17], 9597.
- [19] Xingjie Huang, Jing Li, Jimeng Zhao, Beibei Su, Zixian Dong and Jing Zhang, "Research on Automatic Intrusion Detection Method of Software-Defined Security Services in Cloud Environment" *International Journal of Advanced Computer Science and Applications*[IJACSA], 14[4], 2023. <http://dx.doi.org/10.14569/IJACSA.2023.0140406>
- [20] Agrawal, N., & Kumar, R. [2022]. Security perspective analysis of industrial cyber physical systems [I-CPS]: A decade-wide survey. *ISA transactions*, 130, 10-24.
- [21] Yu, Z., Gao, H., Cong, X., Wu, N., & Song, H. H. [2023]. A Survey on Cyber-Physical Systems Security. *IEEE Internet of Things Journal*.
- [22] Isern, J., Jimenez-Perera, G., Medina-Valdes, L., Chaves, P., Pampiega, D., Ramos, F., & Barranco, F. [2023]. A Cyber-Physical System for integrated remote control and protection of smart grid critical infrastructures. *Journal of Signal Processing Systems*, 1-14.
- [23] Alsakran, F., Bendiab, G., Shiaeles, S., & Kolokotronis, N. [2019, December]. Intrusion detection systems for smart home iot devices: experimental comparison study. In *International Symposium on Security in Computing and Communication* [pp. 87-98]. Singapore: Springer Singapore.
- [24] S. V. B. Rakas, M. D. Stojanović and J. D. Marković-Petrović, "A Review of Research Work on Network-Based SCADA Intrusion Detection Systems," in *IEEE Access*, vol. 8, pp. 93083-93108, 2020, doi: 10.1109/ACCESS.2020.2994961.
- [25] Sangeetha, K., Shitharth, S., & Mohammed, G. B. [2022]. Enhanced SCADA IDS security by using MSOM hybrid unsupervised algorithm. *International Journal of Web-Based Learning and Teaching Technologies* [IJWLT], 17[2], 1-9.
- [26] Potnurwar, A. V., Bongirwar, V. K., Ajani, S., Shelke, N., Dhone, M., & Parati, N. (2023). Deep Learning-Based Rule-Based Feature Selection for Intrusion Detection in Industrial Internet of Things Networks. *International Journal of Intelligent Systems and Applications in Engineering*, 11(10s), 23-35.
- [27] Al-Muntaser, B., Mohamed, M. A., & Tuama, A. Y. (2023). Real-Time Intrusion Detection of Insider Threats in Industrial Control System Workstations Through File Integrity Monitoring. *International Journal of Advanced Computer Science and Applications*, 14(6).
- [28] Mendonça, Y. V., Naranjo, P. G. V., & Pinto, D. C. (2022). The Role of Technology in the Learning Process. *Emerging Science Journal*, 6(Special Issue), 280-295.
- [29] Kandel, I., Castelli, M., & Manzoni, L. (2022). Brightness as an augmentation technique for image classification. *Emerging Science Journal*, 6(4), 881-892.
- [30] Chatterjee, J., & Dethlefs, N. [2020, September]. Temporal causal inference in wind turbine scada data using deep learning for explainable AI. In *Journal of Physics: Conference Series* [Vol. 1618, No. 2, p. 022022]. IOP Publishing.
- [31] Nwakanma, C. I., Ahakonye, L. A. C., Njoku, J. N., Odirichukwu, J. C., Okolie, S. A., Uzundu, C., ... & Kim, D. S. [2023]. Explainable artificial intelligence [xai] for intrusion detection and mitigation in intelligent connected vehicles: A review. *Applied Sciences*, 13[3], 1252.

Model Classification of Fire Weather Index using the SVM-FF Method on Forest Fire in North Sumatra, Indonesia

Darwis Robinson Manalu, Opim Salim Sitompul, Herman Mawengkang, Muhammad Zarlis

Doctoral Study Program (S3) in Computer Science,
Faculty of Computer Science and Information Technology,
Universitas Sumatera Utara, Medan, North Sumatera, Indonesia

Abstract—As a tropical country, Indonesia is situated in Southeast Asia nation has vast forests. Forest fire occur busy vary due to land conditions and forest conditions in drought season. The indicator used mitigated potential forest fire is to study the indicator behavior of the fire weather index (FWI). The data is gathered from the observation station in north Sumatra province, computation and estimation FWI by Canadian Forest Fire Weather Index based on the data gathered. It is found that there is gathered outlier data. to hope will it, it is necessary to conduct classification and predict this of the dataset by machine learning approach using Support Vector Machine Forest Fire (SVM-FF), which is a further development of the previous models, known as the c-SVM and v-SVM. This method includes a balancing parameter by determining the lower and upper limits of a support vector. Furthermore, it allowed the balancing parameter value to be negative. The results showed that the classification of FWI was at low, medium, high, and extreme levels. The low FWI value has an average of 0.5 which is in the 0 to 1 interval. There was an increase in the model's accuracy and performance from its predecessor, which include the c-SVM and v-SVM with respective values of 0.96 and 0.89. Meanwhile, it was observed that with the SVM-FF model, the accuracy was quite better with a value of 0.99, indicating that it is useful as an alternative to classify and predict forest fires.

Keywords—Fire weather index; forest fire; support vector machine; SVM-FF model

I. INTRODUCTION

Forest fire mitigation has been a priority for everyone in order to avoid increasing damage to nature [1]. Several methods applied by forest managers in this process include creating awareness on the prohibition of forest burning and monitoring of fire-prone areas, both of which are man-made and natural factors [2]. One of the data sources used to determine forest fires was the distribution of hotspots [3][4].

A previous research on machine learning that employed a random forest model [5] includes performance measurement of Forest Fire Prediction [6][7]. Another one was the application of machine learning for classification and prediction using fire weather index data[8]. Classification by using multiple variables is a frequently encountered related to data mining problem[9]. The components of the fire weather index in the meteorological data [10] are temperature, rain, wind, air humidity, and other supporting elements for calculating the

Fire Weather Index (FWI) [11] daily in forest areas having fire outbreak potential.

This data source was processed to determine the distribution of FWI in North Sumatra Province. However, the observation data used still needs to be pre-processed to avoid missing value or outlier data [12]. It is important to note that one of the characteristics of meteorological and weather data is the outlier, which are the emergence of extreme parameter values [13]. This makes it to have an impact on model misclassification, biases in parameter estimations, incorrect results, and imperfect forecasts [12] [14], which are classified as a loss function using the Support Vector Machine (SVM) method in machine learning [15][16][17].

The research objective was to classify and predict forest fires in North Sumatra using the fire weather index behavior[18]. The benefit of this research is to provide information on fire weather index values as a reference in predicting the potential for forest fires in an area [19], especially in North Sumatra, and early information on disaster mitigation in the forest fire sector. Research urgency is that the classification of potential forest fires [20] should not be based on the distribution of hotspots alone but needs to be reviewed from the behavior of the fire weather index with the Support Vector Machine Forest Fire (SVM-FF) model approach in reducing and mitigating forest fire disasters in the province of North Sumatra.

II. RELATED WORKS

This research dealing with forest fire classification covers the way how to optimize parameters at SVM to reduce misclassification [21]. With the Weather Research and Forecasting (WRF) mesoscale model method, FWI classification and mapping were carried out in Greece [22].

Further research uses a multi-factor forest fire prediction model with a machine learning approach using the random forest model method which aims to determine the highest incidence in China [23]. The determination of hotspot points that are also used the classification algorithm are C5.0 and Random Forest producing rules-based model[24][25] namely Forest Fire Detection carried out by using an application to determine hotspots in forested areas using Landsat 8 and Band ten satellite images or thermal bands having information on temperature [26][27], Predicting Rainfall from Weather

Observations using SVM Approach to identify The Parameter of Fuel Moisture as Fire Weather Index [28], Predicting fire-prone areas with machine learning techniques [29], and Researching into forest risk assessment system in China [30]. The study presents the classification and prediction of forest fires.

III. FIRE WEATHER INDEX

Obtaining the value of Fire Weather Index (FWI) requires a process that includes meteorological elements. The data collection process starts from observation to data processing [31]. The steps taken include calculating the value of Fire Weather Observations, Fire Behaviors Indices, Fine Fuel Moisture Code (FFMC) [32], Duff Moisture Code (DMC), Drought code (DC), Initial Spread Index (ISI), Buildup Index (BUI), and Fire Weather Index (FWI) [29].

This becomes the basis for processing observation data to produce a fire weather index value in numerical form, which is based on the ISI and BUI, utilized as a general fire hazard index for forest areas [33]. The following equations are employed in the calculation of the FWI value.

$$f(D) = 0.626U^{0.0809} + 2, \quad U \leq 80 \quad (1)$$

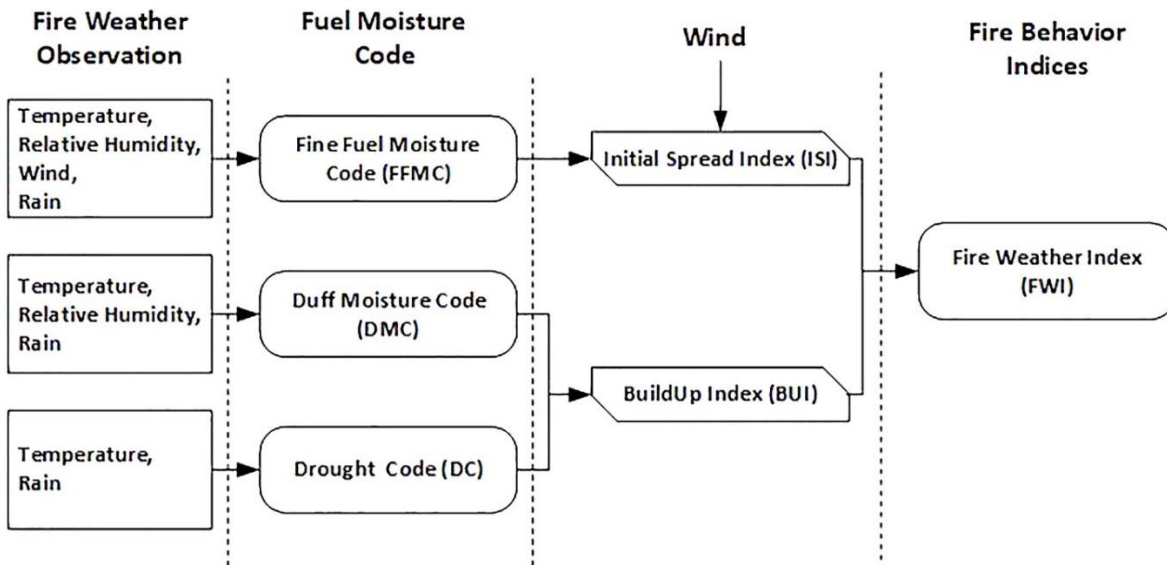


Fig. 1. FWI schematic [34].

One of the classification tasks is building a prediction engine with a good degree of accuracy capable of generalizing [35]. To achieve this purpose, SVM is used to minimize the objective function that contains a loss function [36], with the aim of finding the function $f(x)$ as a hyper-plane and making the error (ϵ) as minor possible [21]. It is important to note that the current studies on SVM generally is focused on improving and formulating loss functions that eventually produce many variants. For example, the C-SVM [37] applied the surrogate function formulation by adding parameters $C \in (0, \infty)$ to the loss function in Eq. (6).

$$\min_{w,b} \frac{C}{2} \|w\|^2 + \sum_{i \in I} [-y_i (w^T x_i + b) + 1]^+ \quad (6)$$

$$f(D) = \frac{1000}{(25+108.64 e^{-0.023U})} + 2, \quad U > 80 \quad (2)$$

$$B = 0.1 R f(D) \quad (3)$$

$$\ln S = 2,72 (0.434 \ln B)^{0.647}, \quad B > 1 \quad (4)$$

$$S = B, Bz \quad (5)$$

Where:

$f(D)$ is the Function of drought (*drought*)

U represents the BUI value

R denotes the ISI value

B the FWI (*intermediate form*)

S represents the FWI (*final form*)

The calculations of these components are based on daily meteorological data observations such as temperature, relative humidity, wind speed, and 24-hour rainfall [32]. The three components of the FWI system provide a numerical rating of the forest's relative fire potential, including Fire Weather Observations, Fuel Moisture Code, and Fire Behaviors Indices. Fig. 1 shows the components of the FWI System.

The v-SVM model in [38] added a parameter to increase the cardinality of the data as expressed in Eq. (7).

$$\min_{w,b,p} v + \frac{1}{|I|} \sum_{i \in I} [-y_i (w^T x_i + b) + 1]^+ \quad (7)$$

In many studies, both C-SVM and v-SVM produced optimal solutions because the parameter v was able to determine the lower and upper bounds of a support vector.

IV. METHODOLOGY

The meteorological data method was used in this research and processing was performed before pre-processing the observation data from all Meteorological, Climatological, and Geophysical Agency (BMKG) stations [39]. The details of the method are shown in Fig. 2.

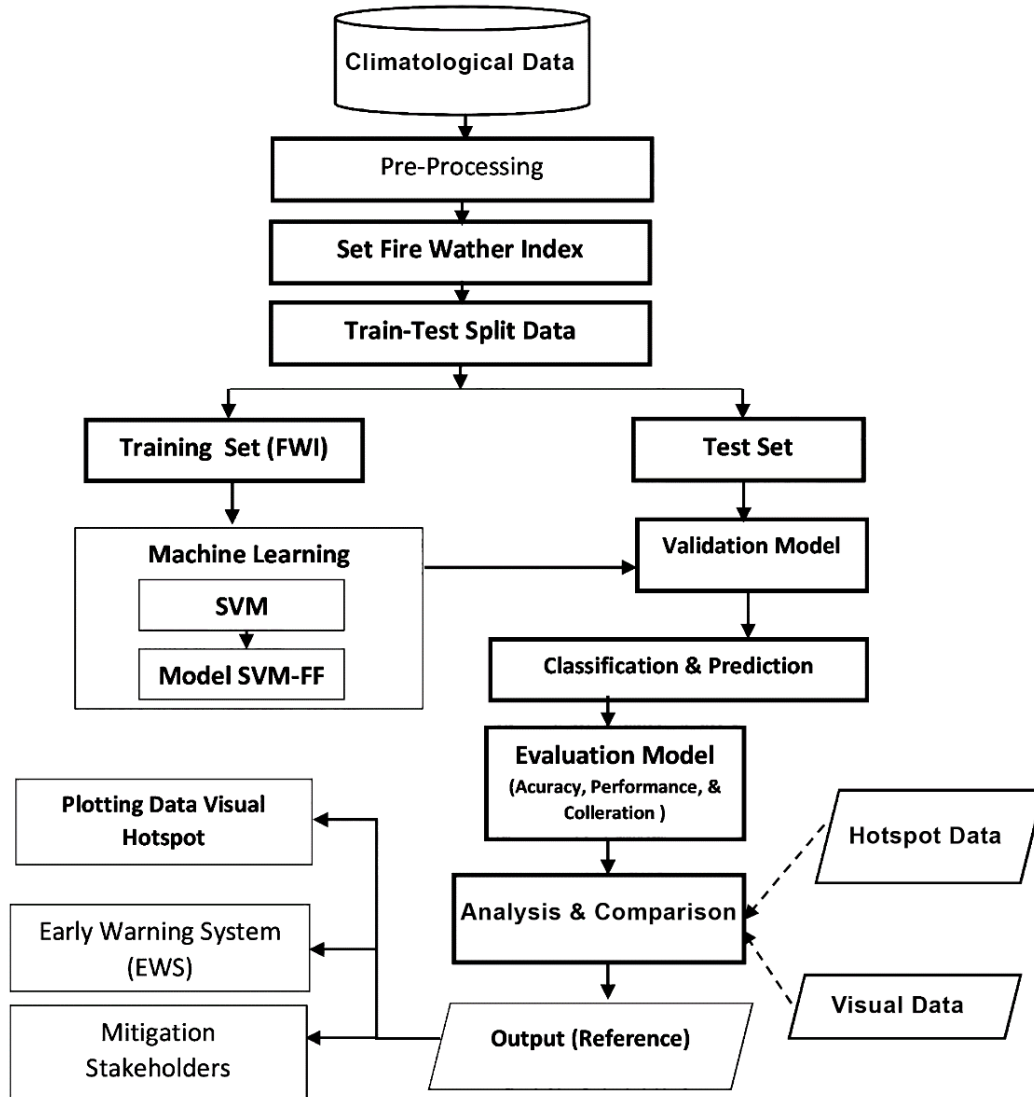


Fig. 2. Research framework.

The SVM-FF focused on increasing negative outlier data in order to get better performance, but when the parameter v in the v -SVM shown in Eq. (7) has an extreme value, the classification model produced poor performance. First, it assigned a large value to the parameter v , which encouraged overfitting [39] and makes the model look too good but not optimal to generalize from the dataset.

Second, it produced a small value of $w=0$ and $b=0$ in Eq. (3) to (5) above, which results in a poor solution. It was observed that the c -SVM and v -SVM models produced optimal solutions because the parameter v was able to determine the lower and upper bounds of a support vector. Based on the above considerations, the proposed solution to force the value of the parameter w in Eq. (7) was to set the value of the parameter p to be negative. This is consistent with [39] that applied optimization to produce negative values as expressed in Eq. (8).

$$C = \min_{w,b,p} \{-vp + \frac{1}{|I|} \sum_{i \in I} [-y_i (w^T x_i + b) + 1]^+\} \quad (8)$$

w : weight

b : biases always positive

p : balancing parameter

y : target/class/category

x : feature data

t : time epoch

i : index

A. SVM-FF Algorithm

The SVM-FF model is based on Eq. (7) with the following steps:

Algorithm 1: SVM-FF model

Initiation value $v \in (0, 1)$, value p, w, b

Minimizing the loss function with equation (8):

$$C = \min_{w,b,p} \left\{ -vp + \frac{1}{|I|} \sum_{i \in I} [-y_i (w^T x_i + b) + 1]^+ \right\}$$

If the value is infinite, the algorithm stops.

Otherwise, the optimal solution is (w_i, b_i)

If value $C_{i+1} = C_i$ the algorithm stops.

Determine as much C_{i+1} sample data from data x_i

Set index to $I_i + 1$

$i = i + 1$

B. Model Validation

The performance test of the SVM-FF model was based on several criteria, such as the average error value using the Receiver Over Characteristic (ROC) curve function [40].

C. Error Plot

An error plot is a graphical representation that shows the range of the actual value to those predicted by the model. When there are many N data observed, then the average of the data is described in Eq. (9) [39].

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (9)$$

According to Eq. (9), the error value is defined as the square root of the difference between the data x_i and the average data \bar{x} as expressed in Eq. (10).

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (10)$$

D. Receiver Operating Characteristic

The most recommended test criterion for multi-class classification was Receiver Operating Characteristic (ROC) analysis [41]. The ROC curve is capable of displaying a sensitivity plot in the y-axis, and 1-specificity in the x-axis. Fig. 3 shows a hypothetical ROC curve that represents diagnostic accuracy. In line A, the curve showed 100% accuracy or 1.0, while in line B, it represents 85% accuracy or 0.85, and curve C depicted 50% accuracy or 0.5.

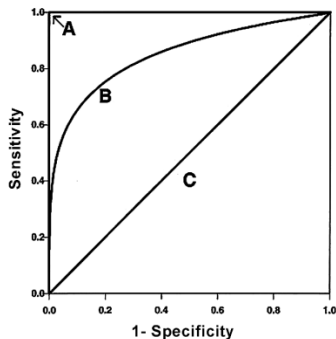


Fig. 3. ROC curve.

V. DISCUSSION

A. BMKG Dataset Exploration

The FWI is a multi-dimensional dataset that consists of 12,897 raw data with ten features and one class. The description of each feature is described in Table I below.

TABLE I. FWI DATASET FEATURES

Feature	Description	Value (max, min, average)
Tn	Real data, minimum temperature	33, 1, 23.25
Tx	Real data, maximum temperature	233, 3, 31.16
Tavg	Real data, mean temperature	31.6, 18.6, 26.9
RH_avg	Real data, average humidity	121, 7, 85.62
RR	Real data, rainfall	8888, 0, 10.23
ss	Real data, duration of sunshine	23, 0, 4.50
ff_x	Real data, maximum wind speed	45, 0, 4.40
ddd_x	Real data, wind direction at maximum speed	2860, 0, 191.62
ff_avg	Real data, average wind speed	15, 0, 1.38
S	Nominal data, dataset class target value	(1,2,3,4)

B. Data Preparation

Climatological dataset contains about 28% of missing value. Data pre-processing involves a series of data preparation process used to handle missing value. Columns in the dataset which are having missing values replaced with the mean of remaining values in the column [42].

C. Condition of Data Outliers

One of the characteristics of disaster, weather, and climate is outlier data. It has been observed that the existence of this data increased as a result of extreme parameters, such as temperature, wind speed, etc. The BMKG data applied in this research was not exempted from the outlier problems. The existence of outliers in data features are shown in Fig. 4 and 5.

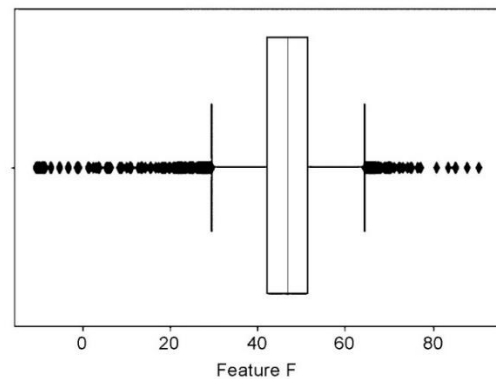


Fig. 4. Outliers on the RH_avg feature (Average Humidity).

Fig. 4 shows the condition of the dataset outliers, specifically for the RH_avg feature or mean humidity. It was observed that some values deviate significantly from the median or the middle value of the data (blue). Furthermore, the outlier values move to the right and left of the median.

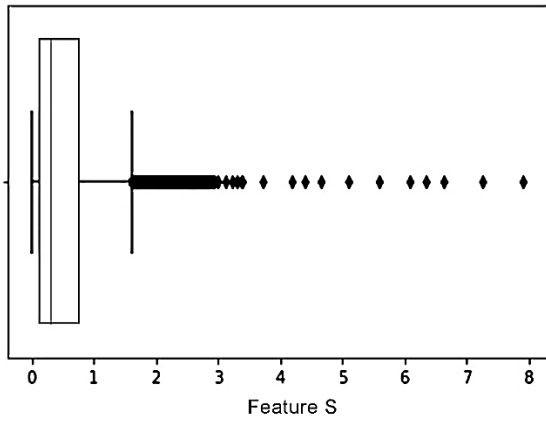


Fig. 5. Outliers on the ff_avg Feature (Average Wind Speed).

The condition of the outlier dataset in Fig. 5 for other features was ff_avg or average wind speed. Furthermore, the outliers are to the right of the data or values above the median. It is important to note that the first step in the classification process with SVM-FF are parameters initiation that include the conventional SVM, namely the kernel type parameter, w value, gamma value, and r coefficient. Other parameters adapted from the SVM-FF model are the values of v and p which have been described in Eq. (8). The description of each feature is described in Table II below.

TABLE II. INITIALIZATION OF SVM-FF PARAMETERS

No	Parameter	Default value	Model Type
1	Kernel	RBF	SVM
2	w	1	SVM
3	Gamma	1	SVM
4	r Coefficient	0	SVM
5	v	(0,1)	SVM-FF
6	p	(0,1)	SVM-FF

The pre-processed FWI data consists of 12,800 raw data, which consist of low, medium, high, and extreme classes. Furthermore, the number of raw data per class was 6,126, 3,394, 3,150 and 130 data, respectively and the graph of these distributions is shown in Fig. 6.

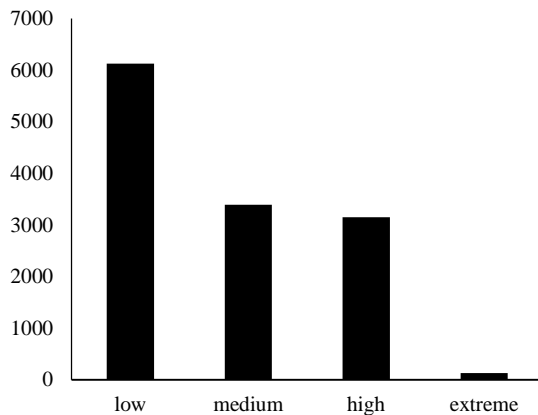


Fig. 6. FWI dataset class distribution graph.

It is important to note that the data validation scheme is based on the k-Fold cross-validation method with $k=10$. This method randomly divided the dataset into ten sub-datasets. In each fold, the training data was nine sub-datasets, while the testing data was one sub-dataset. The distribution scheme of these datasets is shown in Table III below, is showing that in the first fold, the amount of training data was 11,520 and the testing data was 1,280. The same number of data applies from the Second fold through to the tenth fold.

In the testing phase, the SVM-FF model was applied to classify the dataset based on the attribute values of the FWI dataset. The dataset classification results in the first fold (fold-0) are shown in Table III. A total of 25 raw data are taken in both upper data (1 to 25) and lower data (1,256 to 1,280) out of 1,280 raw data in fold-0, respectively.

The data with serial number 1 has an actual class of "medium", and the classification result is "high", making, the classification status was misclassified. Furthermore, the data with sequence number 2 has a "low" actual class and classification result of "low" respectively, which means a correct classification. It can be concluded that in the fold-0, out of 1,280 raw data, 141 data are misclassified and 1,139 with correct status, indicating that the accuracy of data classification in fold-0 is 0.890 globally.

TABLE III. RESULTS OF DATA CLASSIFICATION USING THE SVM-FF MODEL

No	Actual	Classification	Status
1	medium	high	misclassified
2	low	low	correct
3	high	high	correct
4	medium	medium	correct
5	low	low	correct
6	low	low	correct
7	low	low	misclassified
8	medium	low	correct
9	high	high	correct
10	medium	medium	correct
1,272	low	low	correct
1,273	high	high	correct
1,274	extreme	extreme	correct
1,275	high	high	correct
1,276	extreme	high	misclassified
1,277	high	high	correct
1,278	high	high	correct
1,279	extreme	extreme	correct
1,280	medium	low	correct

Similarly, the accuracy of each fold in both the testing and training stages is described in Table IV. It was observed that in fold-0, the training and testing accuracies were 0.978 and 0.890, respectively. When the overall fold was tested with 10 folds, the average training and testing accuracies were 0.983 and 0.906, respectively.

TABLE IV. AVERAGE ACCURACY OF TRAINING AND TESTING

Fold	Testing	Training
Fold 0	0.890	0.978
Fold 1	0.919	0.978
Fold 2	0.957	0.981
Fold 3	0.906	0.987
Fold 4	0.919	0.982
Fold 5	0.894	0.981
Fold 6	0.904	0.982
Fold 7	0.904	0.988
Fold 8	0.879	0.980
Fold 9	0.879	0.988
Correctly classified	0.906	0.983
Global Classification Error	0.154	0.064
Stddev Global Classification Error	0.020	0.003

In addition to the analysis related to the testing phase, further evaluation was performed to determine the performance of each fold classification per class in the training and testing phases. The results are shown in Table V and Table VI, in which the SVM-FF model has the highest accuracy for the "high" class classification and the lowest results for the "medium".

TABLE V. RESULTS OF PERFORMANCE TEST PER FOLD PER CLASS IN THE TRAINING PHASE

	Low	Medium	High	Extreme
Fold0	1.000	0.887	1.000	0.843
Fold1	1.000	0.867	1.000	0.874
Fold2	0.994	0.908	1.000	0.859
Fold3	1.000	0.887	0.995	0.864
Fold4	0.994	0.882	1.000	0.853
Fold5	0.994	0.892	0.995	0.858
Fold6	0.994	0.893	1.000	0.859
Fold7	1.000	0.908	1.000	0.859
Fold8	1.000	0.872	1.000	0.864
Fold9	1.000	0.897	1.000	0.869
Average	0.998	0.889	0.999	0.860

TABLE VI. RESULTS OF PERFORMANCE TEST PER FOLD PER CLASS IN THE TESTING PHASE

	Low	Medium	High	Extreme
Fold-0	0.880	0.838	0.965	0.692
Fold-1	0.950	0.727	1.000	0.905
Fold-2	0.950	0.818	1.000	0.619
Fold-3	0.950	0.864	0.909	0.714
Fold-4	1.000	0.727	0.952	0.682
Fold-5	0.950	0.636	0.952	0.818
Fold-6	0.950	0.714	1.000	0.714
Fold-7	0.950	0.810	0.955	0.667
Fold-8	0.950	0.714	0.955	0.667
Fold-9	1.000	0.545	0.955	0.810
Average	0.953	0.739	0.964	0.729

D. SVMFF Model Performance

The performance test of the SVM-FF model is based on several criteria, which includes the average value of the error and the Receiver Over Characteristic (ROC) Curve.

E. Error Graph

The error plot of a graph property represents variations in the data, which is due to the errors or uncertainty of a model in the form of visualization of vertical lines in the error point. The benchmark for determining the error plot was Standard Deviation, and the graph for that of SVM-FF model was visualized in Fig. 7 and Fig. 8.

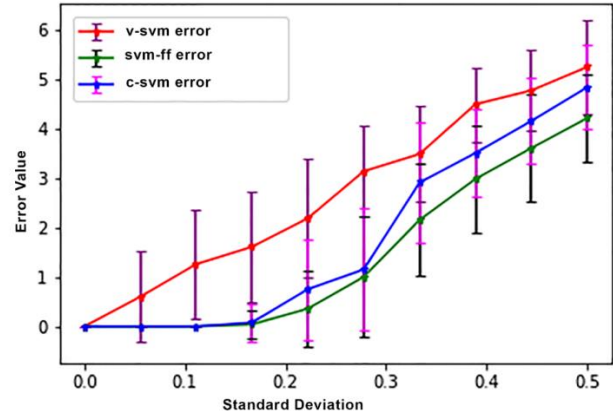


Fig. 7. C-SVM, v-SVM, and SVM-FF model error graphs.

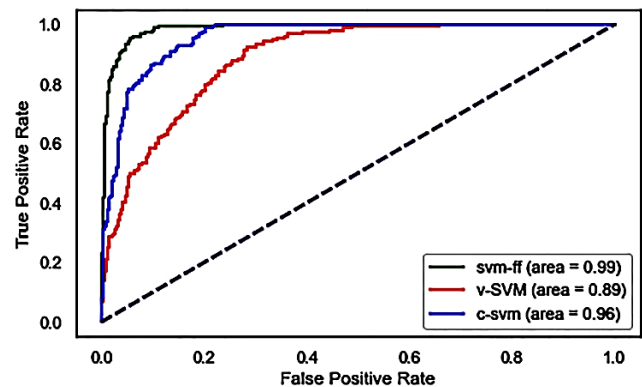


Fig. 8. ROC Comparison graph between SVM-FF, v-SVM, and c-SVM.

It was observed that the error graph or error plot for the SVM-FF model was lower or better compared to the other two models, which include c-SVM and v-SVM.

F. ROC Curve

The ROC curve shows the probability or accuracy of the model in graphical form. According to the test results, the SVMFF model has a better model accuracy in comparison with the other two SVM.

The ROC graph above shows the accuracy and performance of the model, in which the c-SVM and v-SVM models produced accuracies of 0.96 0.89, respectively. When the SVM-FF model was used, the accuracy was improved with a value of 0.99.

G. Decision Region Boundary (DRB)

A two-dimensional graph represents the correspondence of multiple data feature to classes. It is important to note that the SVM-FF method has three types of kernels, which include linear, sigmoid, and Radial Basis Function (RBF). The application of these three kernels affected the shape of the SVM-FF hyperplane. The SVM-FF DRB model is shown in Fig. 9, 10, and 11.

There are four classes of fire weather index datasets, comprising low, medium, high, and extreme respectively indicated with blue, orange, green, and red. The classification was performed in the SVM-FF by placing data features in different regions. It was observed that when the linear kernels were used, there were still deviations in the classification areas,

particularly in the extreme classes, which are marked with red circles.

The data area also appears to have been classified, but some data points still occupied different positions. For example, the maximum margin is still very close to the hyperplane in the form of a straight line. Meanwhile, in the data section with a yellow triangle symbol, some still occupy the blue and green areas. This shows that the accuracy with the kernel liner was not optimal as expected.

Furthermore, with the use of the sigmoid kernel, the position of the hyperplane was observed after the data points. The sigmoid line shape in the green round category is seen in Fig. 11. Based on the three figures and the classification results using the kernel, the RBF type presented a better class distinction when compared to linear or sigmoid kernels.

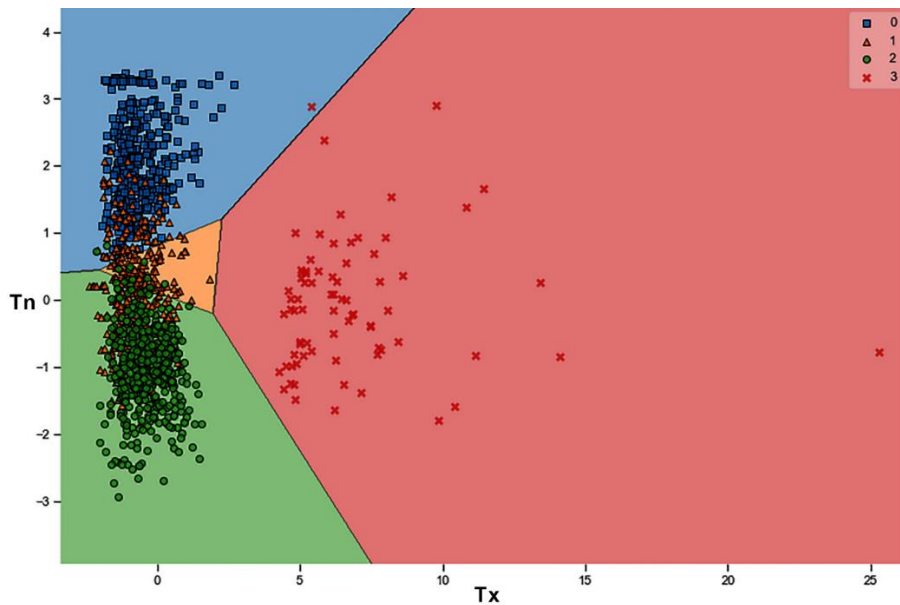


Fig. 9. DRB with linear kernel.

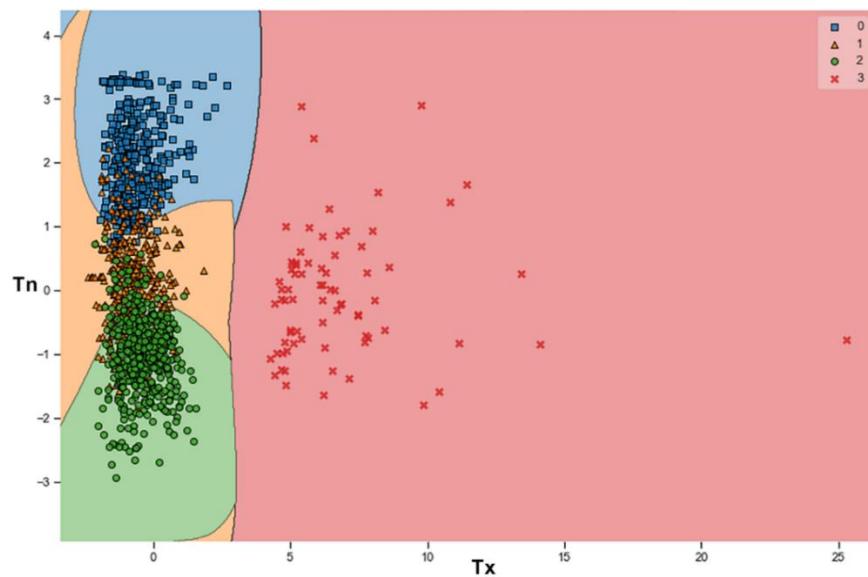


Fig. 10. DRB with RBF kernel.

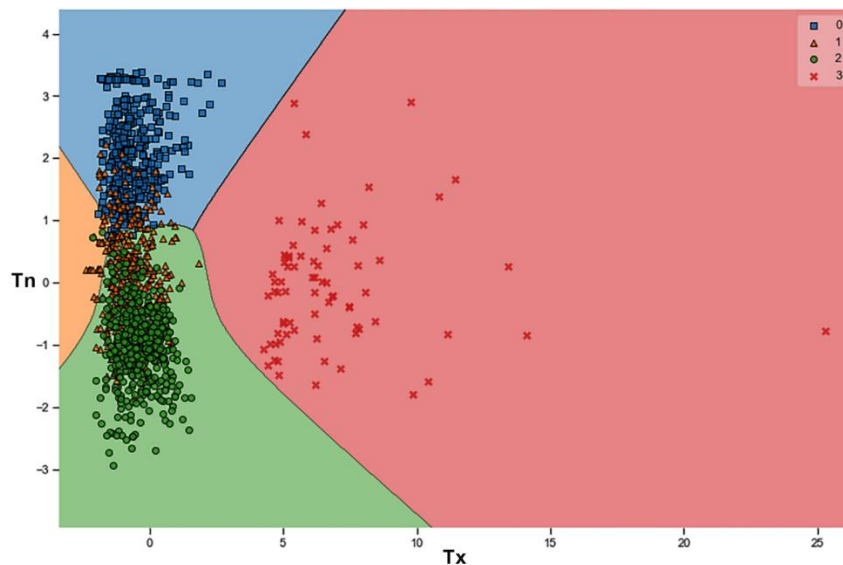


Fig. 11. DRB with sigmoid kernel.

H. Model Significance Test

A statistical significance test approach was used to test the significance of the models as shown in Table VII. The Wilcoxon test was applied at a significance level of 95% or an error rate of 5% (0.05), and was declared to be significantly different when the p-value is < 0.05 .

TABLE VII. THE WILCOXON TEST RESULTS OF SVM-FF MODEL

VS	R^+	R^-	Exact P-value	Asymptotic P-value
v-svm	55.0	0.0	0.00195	0.004317
c-svm	55.0	0.0	0.00195	0.004317

According to Table VII, the p-value of SVM-FF VS v-SVM was $0.00195 < 0.05$, indicating that the accuracy of SVM-FF differs significantly from the v-SVM. When the p-value of SVM-FF VS c-SVM was $0.00195 < 0.05$, the SVM-FF accuracy was observed to differ significantly from the c-SVM.

VI. CONCLUSION

In conclusion, the testing results when the SVM-FF model was used for FWI conditions in North Sumatra Province provided four classifications, namely low, medium, high, and extreme, indicated as blue, orange, green, and red zones, respectively. From 2017 to 2020, FWI conditions have been classified as a low-level/blue zone with an average value of 0.5 per day, which is on a scale of 0 and 1.

This classification was conducted using all observational datasets by entering outlier data to determine the performance optimization of the SVM-FF model and was compared to the result of its predecessor. It was observed that the c-SVM and v-SVM models produced accuracies of 0.96 and 0.89, respectively. Meanwhile, the SVM model -FF was able to improve the accuracy with a value of 0.99, indicating that it was able to work more optimally.

The model's performance when several kernel functions were utilized showed that the Radial Basis Function kernel provided classification results with a better hyper-plane

position compared to the maximum position of the data margin.

REFERENCES

- [1] M. B. R. Prayoga and R. H. Koestoer, "Improving Forest Fire Mitigation in Indonesia: A Lesson from Canada," *J. Wil. dan Lingkungan.*, vol. 9, no. 3, pp. 293–305, 2021, doi: 10.14710/jwl.9.3.293-305.
- [2] M. Batista, B. Oliveira, P. Chaves, J. C. Ferreira, and T. Brandao, "Improved Real-time Wildfire Detection using a Surveillance System," *Lect. Notes Eng. Comput. Sci.*, vol. 2240, pp. 520–526, 2019.
- [3] Z. Muchamad and M. Emanuel, "Classification of Hotspots Causing Forest and Land Fires Using the Naive Bayes Algorithm," pp. 555–567, 2022.
- [4] M. Zainul and E. Minggu, "Classification of Hotspots Causing Forest and Land Fires Using the Naive Bayes Algorithm," *Interdiscip. Soc. Stud.*, vol. 1, no. 5, pp. 555–567, 2022, doi: 10.55324/iss.v1i5.62.
- [5] A. Apostolakis, S. Girtsou, C. Kontoes, I. Papoutsis, and M. Tsoutsos, "Implementation of a Random Forest Classifier to Examine Wildfire Predictive Modelling in Greece Using Diachronically Collected Fire Occurrence and Fire Mapping Data," vol. 12573 LNCS, no. March. Springer International Publishing, 2021. doi: 10.1007/978-3-030-67835-7_27.
- [6] Singh, K. P. Neethu, K. Madhurekaa, A. Harita, and P. Mohan, "Parallel SVM model for forest fire prediction," *Soft Comput. Lett.*, vol. 3, no. July, p. 100014, 2021, doi: 10.1016/j.socl.2021.100014.
- [7] H. Y. Lin, "Effective feature selection for multi-class classification models," *Lect. Notes Eng. Comput. Sci.*, vol. 3 LNECS, pp. 1474–1479, 2013.
- [8] H. Miyajima, H. Miyajima, and N. Shiratori, "Fast and secure edge-computing algorithms for classification problems," *IAENG Int. J. Comput. Sci.*, vol. 46, no. 4, pp. 1–6, 2019.
- [9] H. Y. Lin and Y. H. Lai, "Construction and evaluation of a robust classification model for multi-objective problems," *IMECS 2011 - Int. MultiConference Eng. Comput. Sci. 2011*, vol. 1, pp. 346–350, 2011.
- [10] BMKG, "Bddan Metereologi, Klimatologi dan Geofisika," BMKG, 2021. <https://www.bmkg.go.id/cuaca/prakiraan-cuaca-indonesia.bmkg?Prov=34&NamaProv=Sumatera Utara> (accessed Jun. 17, 2021).
- [11] CWFIS, "Background Information Canadian Forest Fire Weather Index (FWI) System," 2017. [Online]. Available: <https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi>
- [12] L. Sunitha, D. M. BalRaju, Kiran, and J. Sasi, "Detection and Analysis of Outliers and Applying Data Mining Methods on Weather Data of Bhanur

- Village Detection and Analysis of Outliers and Applying Data Mining Methods on Weather Data of Bhanur Village Abstract.," no. January, 2021.
- [13] C. an Hsiao and H. Chen, "On classification from the view of outliers," *IAENG Int. J. Comput. Sci.*, vol. 37, no. 4, 2010.
- [14] D. R. Manalu, M. Zarlis, H. Mawengkang, and O. S. Sitompul, "Forest Fire Prediction in Northern Sumatera using Support Vector Machine Based on the Fire Weather Index," *AIRCC Publ. Corp.*, vol. 10, no. 19, pp. 187–196, 2020, doi: 10.5121/csit.2020.101915.
- [15] V. Kecman, "Support Vector Machines – An Introduction," no. May 2005, pp. 1–47, 2005, doi: 10.1007/10984697_1.
- [16] S. S. Mehta and N. S. Lingayat, "Support vector machine for cardiac beat detection in single lead electrocardiogram," *Lect. Notes Eng. Comput. Sci.*, vol. 2, no. May, pp. 1630–1635, 2007.
- [17] Z. J. Lv, Q. Xiang, and J. G. Yang, "Application of Genetic Algorithm-support vector machine for prediction of spinning quality," *Proc. World Congr. Eng. 2011, WCE 2011*, vol. 2, pp. 1033–1038, 2011.
- [18] N. Members, "Fire Weather Index (FWI) System," National Wildfire Coordinating Group, 2021. <https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system> (accessed Apr. 21, 2021).
- [19] N. Coordinating Group Wildfire, "Fire Weather Index (FWI) System," NWCG, 2022. <https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system>
- [20] M. Yandouzi et al., "Forest Fires Detection using Deep Transfer Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 8, pp. 268–275, 2022, doi: 10.14569/IJACSA.2022.0130832.
- [21] A. Tharwat, A. E. Hassanien, and B. E. Elnaghi, "A BA-based algorithm for parameter optimization of Support Vector Machine," *Pattern Recognit. Lett.*, vol. 93, pp. 13–22, 2017, doi: <https://doi.org/10.1016/j.patrec.2016.10.007>.
- [22] Varela, "Fire Weather Index (FWI) classification for fire danger assessment applied in Greece," *Tethys, J. Weather Clim. West. Mediterranean*, no. 1994, pp. 31–40, 2015, doi: 10.3369/tethys.2018.15.03.
- [23] L. Yudong, F. Zhongke, Z. Ziyu, C. Shilin, and Z. Hanyue, "Research on Multi-Factor Forest Fire Prediction Model Using Machine Learning Method in China," 2020.
- [24] G. P. Siknun and I. S. Sitanggang, "Web-based classification application for forest fire data using the shiny framework and the C5.0 algorithm," *Procedia Environ. Sci.*, vol. 33, pp. 332–339, 2016, doi: 10.1016/j.proenv.2016.03.084.
- [25] G. ElSharkawy, Y. Helmy, and E. Yehia, "Employability Prediction of Information Technology Graduates using Machine Learning Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, pp. 359–367, 2022, doi: 10.14569/IJACSA.2022.0131043.
- [26] B. Tien, Dieu, H. N. Duc, and S. Pijush, "Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization: A case study at Lao Cai province (Viet Nam)," *J. Environ. Manage.*, vol. 237, no. January, pp. 476–487, 2019, doi: 10.1016/j.jenvman.2019.01.108.
- [27] A. Srivastava, S. Umrao, S. Biswas, and I. Zafar, "FCCC : Forest Cover Change Calculator User Interface for Identifying Fire Incidents in Forest Region using Satellite Data," vol. 14, no. 7, 2023.
- [28] Wijayanto, S. O. K. N D, and H. Y, "Classification Model for Forest Fire Hotspot Occurrences Prediction Using ANFIS Algorithm," *IOP Conf. Ser. Earth Environ. Sci.* 54 012059, no. January, 2017, doi: 10.1088/1755-1315/54/1/012059.
- [29] A. M. Elshewey, "Forest Fires Detection Using Machine Learning Techniques," vol. XII, no. IX, pp. 510–517, 2020, [Online]. Available: <https://www.xajzkjdx.cn/gallery/54-sep2020.pdf>
- [30] X. Chen, T. Li, L. Ruan, K. Xu, J. Huang, and Y. Xiong, "Research and application of fire risk assessment based on satellite remote sensing for transmission line," *Lect. Notes Eng. Comput. Sci.*, vol. 2219, pp. 284–287, 2015.
- [31] D. R. Manalu, M. Zarlis, H. Mawengkang, and O. S. Sitompul, "Predicting rainfall from weather observations using SVM approach for identify the parameter of fuel moisture code as fire weather index," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 16, pp. 4090–4097, 2021.
- [32] Canada.ca, "Natural Resources Canada," Government of Canada, 2020. <https://cwffis.cfs.nrcan.gc.ca/background/summary/fwi>
- [33] M. Noor, Kebakaran Lahan Gambut. Universitas Lambung Mangkurat, 2019. [Online]. Available: <http://eprints.ulm.ac.id/9594/1/2>. Kebakaran Lahan Gambut-Faktor Penyebab dan Mitigasinya.pdf
- [34] G. of Canada, "Canadian Wildland Fire Information System | Canadian Forest Fire Weather Index (FWI) System," <https://cwffis.cfs.nrcan.gc.ca/background/summary/fwi>, 2020. <https://cwffis.cfs.nrcan.gc.ca/background/summary/fwi> (accessed Oct. 19, 2020).
- [35] S. D. Jena, J. Kaur, Rani, and Rajneesh, "A Review of Prediction of Software Defect by Using Machine Learning Algorithms BT - Recent Innovations in Computing," 2022, pp. 61–70.
- [36] M. Tanveer, A. Sharma, and P. N. Suganthan, "General twin support vector machine with pinball loss function," *Inf. Sci. (Ny.)*, vol. 494, pp. 311–327, 2019, doi: <https://doi.org/10.1016/j.ins.2019.04.032>.
- [37] I. Ibrahim, R. Silva, M. H. Mohammadi, V. Ghorbanian, and D. A. Lowther, "Surrogate-Based Acoustic Noise Prediction of Electric Motors," *IEEE Trans. Magn.*, vol. 56, no. 2, pp. 1–4, 2020, doi: 10.1109/TMAG.2019.2945407.
- [38] Y. Wang, X. Tang, H. Chen, T. Yuan, Y. Chen, and H. Li, "Sparse additive machine with pinball loss," *Neurocomputing*, vol. 439, pp. 281–293, 2021, doi: <https://doi.org/10.1016/j.neucom.2020.12.129>.
- [39] BMKG, "BMKG, Data Online Pusat Database," 2020. <https://dataonline.bmkg.go.id/home>
- [40] P. A. Emelia Akashah, S. K. Sugathan, and A. T. S. Ho, "Receiver operating characteristic (ROC) graph to determine the most suitable pairs analysis threshold value," *Proc. - Adv. Electr. Electron. Eng. - IAENG Spec. Ed. World Congr. Eng. Comput. Sci. 2008, WCECS 2008*, no. November, pp. 224–230, 2008, doi: 10.1109/WCECS.2008.35.
- [41] K. H. Zou, O'Malley, A. James, and L. Mauri, "Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models," *Circulation*, vol. 115, no. 5, pp. 654–657, 2007, doi: 10.1161/CIRCULATIONAHA.105.594929.
- [42] N. Mamat, S. Fatim, and M. Razali, "Comparisons of Various Imputation Methods for Incomplete Water Quality Data: A Case Study of The Langat River , Malaysia," vol. 35, no. 1, pp. 191–201, 2023.

Prediction of Cryptocurrency Price using Time Series Data and Deep Learning Algorithms

Michael Nair¹, Mohamed I. Marie², Laila A. Abd-Elmegid³

Department of Management Information Systems, Higher Institute of Qualitative Studies, Heliopolis, Cairo, Egypt¹
Associate professor, Department of Information Systems-Faculty of Computers and Artificial Intelligence,
Helwan University, Cairo, Egypt^{2,3}

Abstract—One of the most significant and extensively utilized cryptocurrencies is Bitcoin (BTC). It is used in many different financial and business activities. Forecasting cryptocurrency prices are crucial for investors and academics in this industry because of the frequent volatility in the price of this currency. However, because of the nonlinearity of the cryptocurrency market, it is challenging to evaluate the unique character of time-series data, which makes it impossible to provide accurate price forecasts. Predicting cryptocurrency prices has been the subject of several research studies utilizing machine learning (ML) and deep learning (DL) based methods. This research suggests five different DL approaches. To forecast the price of the bitcoin cryptocurrency, recurrent neural networks (RNN), long short-term memories (LSTM), gated recurrent units (GRU), bidirectional long short-term memories (Bi-LSTM), and 1D convolutional neural networks (CONV1D) were used. The experimental findings demonstrate that the LSTM outperformed RNN, GRU, Bi-LSTM, and CONV1D in terms of prediction accuracy using measures such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared score (R^2). With RMSE= 1978.68268, MAE=1537.14424, MSE= 3915185.15068, and R^2 = 0.94383, it may be considered the best method.

Keywords—Cryptocurrency; deep learning; prediction; LSTM

I. INTRODUCTION

The fiat currency used in the present monetary system has various disadvantages, including government control over the money supply; transactions are often carried out via intermediaries like financial institutions, which results in expensive fees and prolonged transfer times, as well as the present ledgers used to record transactions being vulnerable to manipulation [1]. Hence Due to its decentralization, immutability, and security, cryptocurrencies have become a worldwide phenomenon that draws a sizable number of users. They are founded on confidence in technology infrastructure, enabling money transfer from any location with nearly negligible delay [2]. Throughout its limited life, the cryptocurrency market has expanded irrationally and astoundingly [3].

Bitcoin, a kind of electronic money, was originally launched in 2008 and was first used as an open-source in 2009 by a person named Satoshi Nakamoto [4]. As the first currency ever created, it has become the most significant currency [5]. Without a single administrator or central bank, it is decentralized digital money that may be transmitted between

users on a peer-to-peer network without the involvement of mediators like banks [6].

Most cryptocurrencies use blockchain technology and feature attributes like decentralization, transparency, and immutability [7]. Blockchain allows for the permanent recording of network transactions [8], and each record is encrypted and carries the block's [9] cryptographic hash before it. A date, sender and recipient details, and the total amount of money transmitted are all included in each record. An extremely complex technology called a secure shell links transaction blocks [10]. This technology aims to store data that makes it difficult or impossible to alter, hack, or defraud the system [11].

From 2009 to 2017, the price of Bitcoin increased to over USD 20,000. As of December 2019, the daily average market volume was around USD 19.45 billion [12], and as of April 2021, the price of Bitcoin hit an all-time high of around \$65000 [13]. Although investments have yielded rich returns, the constant price swings seen by most cryptocurrencies make them difficult and hazardous [14]. Consequently, it takes work to anticipate the price of cryptocurrencies.

Additionally, the sharp variations in bitcoin prices have emerged as a brand-new worldwide concern. Therefore, it is crucial to foresee changes in the price of Bitcoin [15]. Because of this, investors need a forecasting strategy to efficiently capture swings in the price of cryptocurrencies to reduce risk and boost profits [16].

Cryptocurrency price prediction is a time series prediction problem in its early phases. In contrast, older methods were used to anticipate time series based on linear hypotheses and required information that could be categorized as trend or seasonal [3], such as sales forecasting. Due to the extreme volatility and lack of seasonality in the Bitcoin market, these strategies are unsuccessful. Based on its success in similar domains, deep learning is an attractive technological choice, given the difficulty of the challenge [17]. From this point on, DL methods are considered efficient for time series forecasting since they are noise-resistant, can accommodate data sequences natively, and can recognize non-linear temporal correlations on such sequences [18].

Estimating the price of the Bitcoin cryptocurrency is the aim of this research, and evaluating the forecasting accuracy of five different deep learning models, including LSTM, RNN, GRU, Conv1D, and Bi-LSTM. The research uses the (RMSE),

(MAE), (MSE), and (R^2) as measurement techniques to assess the performance of DL models using the closing price of Bitcoin in USD.

The issue that motivated us to conduct this paper was the lack of a specific model with high accuracy that could be relied upon to predict the price of cryptocurrencies, which may have a significant impact on the increase in financial profits. It was vital to provide a solid approach to address this issue for investors that invest in these encrypted currencies. Deep learning methods were used as a consequence because they produced positive outcomes in various study domains.

This paper contributes to providing knowledge to everyone interested in this field in identifying deep learning techniques and their ability to deal with time series data to predict the prices of cryptocurrencies, where the results of the research proved that the use of deep learning method resulted in better results than the traditional machine learning techniques, and also to assist investors interested in trading cryptocurrencies in selecting the best deep learning model to predict prices, and to make the right decision to decrease their loss exposure and increase profitability during the trading process in this currency.

The paper is divided into seven sections: Section 2 is a literature review, Section 3 provides background knowledge, and Section 4 presents the model to guide our approach. Section 5 tests the suggested model; Section 6 presents the findings of the experiment; Section 7 conclusions and future work.

II. LITERATURE REVIEW

Bitcoin is a cryptocurrency and a kind of electronic money. It is a well-known cryptocurrency with a bright future [19], and it is a web-based trade technique that uses cryptographic tools to carry out financial transactions [20]. It is crucial to forecast the values of this currency because of the considerable price volatility of this encrypted money, which has the potential to impact investors negatively and international and commercial ties [21]. Numerous researches have been carried out to forecast time series and the value of bitcoin [10]. In contrast, deep learning models [13] and machine learning models [4] were employed to forecast the price of Bitcoin.

The prior research on predicting cryptocurrency prices will be examined in the following part, employing various ML and DL models for time series prediction, as shown in Table I.

TABLE I. LITERATURE REVIEW FOR CRYPTOCURRENCY PRICE PREDICTION PRICE USING ML AND DL

Author	Year	Technique	Cryptocurrency	Dataset Source	Data Range	Prediction Methods	Performance Measures and results	Demerit
HASAN et al [7]	2022	DL	Bitcoin Ethereum Monero	Investing.com	between Jan 22, 2015 to Feb 12, 2020	LSTM, RNN and Proposed method	The Proposed method has achieved the best performance when predicting Bitcoin price with MSE= 18.65, MAE= 2.15 and RMSE= 4.21	Not Explored time-series model such as GRU
NEMATALLAH et al [10]	2022	DL	Bitcoin	Kaggle	between 1 Jan 2012 to 31 Mar 2021	RNN LSTM	MAPE and RMSE LSTM performs better than RNN	Not Explored time-series model such as GRU
Bitto et al [22]	2022	ML	Bitcoin, Ethereum, Litecoin and Tether token	Yahoo Finance	between 2015-1-1 to 2021-6-1	AR MA ARMA	MAE and RMSE. AR model giving better performs than others models with 97.21% For bitcoin, 96.04% for Ethereum, 95.8% for Litecoin and 99.91% accuracy for Tether-token	Not considered deep learning models for prediction
Ammer et al [12]	2022	DL	AMP, Ethereum, Electro-Optical System, and XRP	CoinMarketCap	between May 2015 through April 2022	LSTM	MSE, RMSE, NRMSE and R. LSTM achieved R = 96.73% for training And R= 96.09% for testing when predicting XRP price	Not Explored time-series model such as GRU
FAKHARCHIAN et al [15]	2022	DL	Bitcoin	Yahoo Finance	between 05/02/2021 To 10/09/2021	proposed models based on CNN and LSTM	Model-9 achieved the best performance with MSE= 0.00151, RMSE= 0.0388, MAE= 0.02519, MedAE= 0.01747 and R2= 0.98219	Not Explored time-series model such as GRU
ZHANG et al [23]	2022	ML DL	Bitcoin	Data.Bitcoinity. Org, Blockchain.com , and CoinMarketCA P	between 05/02/2021 To 10/09/2021	LSSVM BP SDAE-B	SDAE-B model giving better performs than others models with MAPE= 0.016, RMSE= 131.643 and DA= 0.817	Not Explored time-series model such as LSTM and GRU

GURRIB et al [24]	2022	ML	Bitcoin	CoinMarketCA P	between 17 Jun 2016 to 21 Apr 2021	LDA SVM	LDA model giving better performs than SVM with accuracy of 0.585	Not considered deep learning models for prediction
CAVALLI et al [25]	2021	ML	Bitcoin	CoinMarketCA P	between 28 of Apr 28, 2013 to Feb 15, 2020	1D CNN LSTM	RMSE 1D CNN model giving better performs than LSTM	Not considered deep learning models for prediction
LIU et al [26]	2021	ML DL	Bitcoin	Coindesk.com BTC.com	between Jul 2013 to Dec 2019	BPNN SVR SDAE	SDAE model giving better performs than others models with MAPE= 0.1019, RMSE= 160.63 and DA= 0.5985	Not Explored time-series model such as LSTM and GRU
MARNE et al [27]	2020	ML DL	Bitcoin	Kaggle	between Jan 2014 to Jan 2019	SVM RNN LSTM	LSTM model giving better performs than others models with RMSE = 3.38	Not Explored time-series model such as GRU

III. METHODOLOGY

We will go through several related concepts in this section.

A. Time-Series

It is one of the most effective methods for forecasting situations with some degree of future uncertainty by analyzing past patterns and assuming that future trends will be similar. Time series forecasting is also based on data for efficient and effective planning to solve forecasting problems with a time component [3].

B. Deep Learning Methods used for Bitcoin Price Prediction

The approaches utilized in DL, a subfield of ML, are built on the structure and design of ANNs. Five DL algorithms were used in this study to forecast the price of Bitcoin. LSTM, GRU, BiLSTM, simple RNN, and the 1D CNN algorithm.

1) *Recurrent Neural Network (RNN)*: Artificial neural networks were inspired by how the human brain processes information. The neural network comprises synthetic neurons, and its architecture determines its properties. Traditional neural networks do not have feedback loops, which is how RNNs vary from them. It is thus relevant anytime the input context affects how well a prediction is made. Each neuron's current state depends on its past state due to the recurrent nature of an RNN's layers, which leaves the neural network with a finite amount of memory. Sequential data may be input into a recurrent neural network, and both the networks In and Out may be sequences of variable lengths that pass through each cell consecutively [28]. Suppose there is an input neuron X_t , an invisible output status h_t , and the prior invisible output status h_{t-1} . In that case, the RNN has a single-layer recurrent module with a tanh squashing function. Fig. 1 [29] demonstrates that W represents the weighted matrix and y_t for the result.

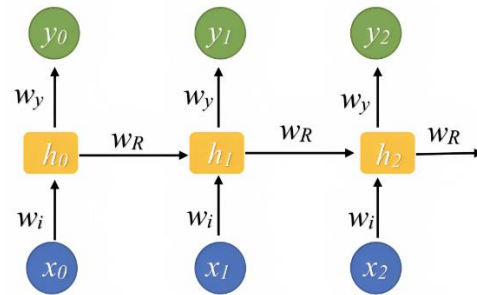


Fig. 1. The recurrent neural network simple architecture [29].

2) *Long Short-Term Memory (LSTM)*: Recurrent neural networks with the ability to learn long term dependencies are called LSTMs. The recommended networks by Hochreiter and Schmid Huber [30] because the last state needed to be sufficiently recent and thus influenced the present state, the RNN model may inaccurately predict the current state [31]. From left to right, the LSTM is crafted to keep track of information throughout time and lessen the issue of vanishing gradient descent. Three interconnected layers in the LSTM, the input gate, forget gate, and output gate, control the data flow necessary to forecast the output of the network [32].

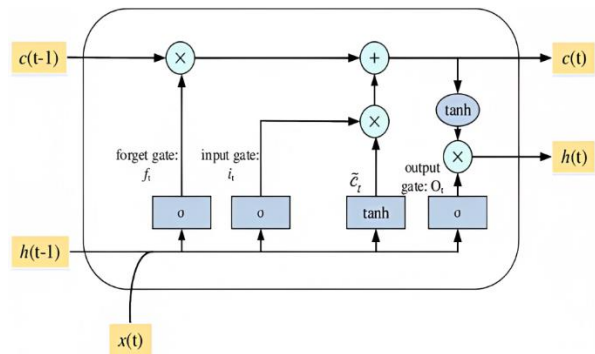


Fig. 2. Schematic diagram of LSTM [33].

Input gate: Information will initially pass through the input gate after importing the data. The switch decides whether or not to store the information based on the state of the cell.

Output gate: The amount of output information is determined by it.

Forget gate: It chooses whether to keep or forget the information obtained. [34, 35], as shown in Fig. 2.

3) *Gate Recurrent Unit (GRU)*: Another RNN version is a GRU, which combines the three gated units into only two gated units: the gate for updating and resetting [36]. GRUs address the vanishing gradient issue of RNNs and the optimization of the structure of the LSTM model. The two gates may store relevant data in the memory cell while transmitting values to the network's later stages. GRU and LSTM are equal when evaluating performance across various test scenarios [37]. Fig. 3 depicts the organization of the GRU units.

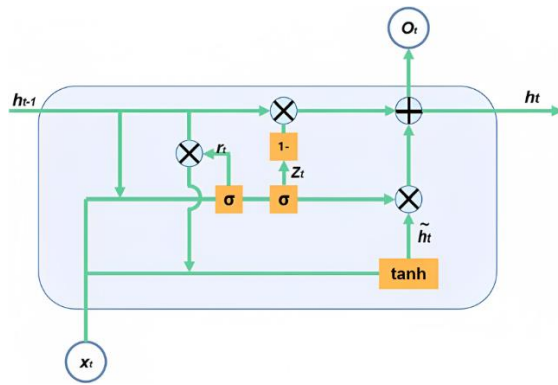


Fig. 3. GRU unit structure [38].

4) *Bidirectional Long Short-Term Memory (BiLSTM)*: The BiLSTM model can extract contextual information from feature sequences by considering both forward and backward dependencies. Using a front LSTM, that processes the sequence in chronological order and a backward LSTM that processes the sequence in reverse order, BiLSTM allows looking ahead. The output is then produced by joining the LSTM's forward and reverse states [39, 40] as seen in Fig. 4.

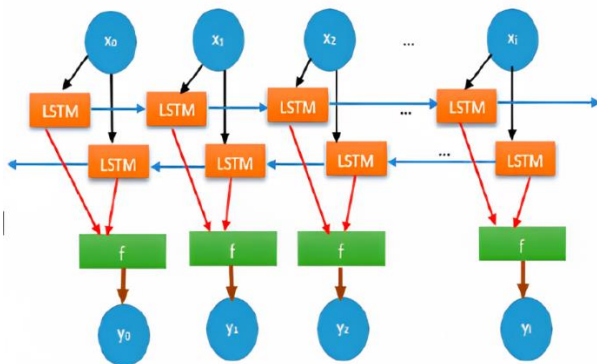


Fig. 4. BiLSTM architecture [11].

5) *1D Convolutional Neural Network (CONVID) model*: It is easy to find basic patterns in data using a convolutional neural network (CNN), which is then used to build more complex patterns in the top layers. A 1D CNN is helpful when

extracting key features from tiny (fixed-length) segments of the whole dataset. The feature's location within the segment is irrelevant; this is correct for analyzing historical data and evaluating sensor data time series. Input, output, and hidden layers comprise a CNN; a feedforward neural network is created using the intermediary layers. Since their inputs and outputs are blind to the activation function and final convolution [31], as illustrated in Fig. 5, these are called hidden layers.

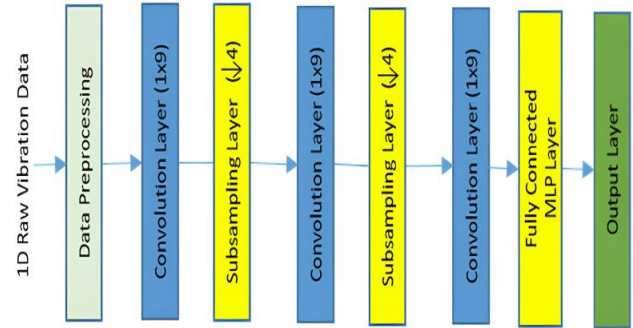


Fig. 5. 1D CNN architecture [41].

IV. THE PROPOSED MODEL FOR PREDICTING CRYPTOCURRENCY PRICE MOVEMENT

This section's suggested model focuses on three key elements. Fig. 6 illustrates the three steps used to anticipate the movement of the cryptocurrency price: (1) Dataset; (2) Data pre-processing; and (3) Deep learning-based algorithms.

Table I displays the literature review, methods used, and limitations of each study, which show the inaccuracy, the use of primitive methods, or a small dataset are all examples of shortcomings. In our research, we used similar and different methods, such as RNNs, LSTMs, GRUs, CONV1D, Bi-LSTM, and different datasets with large sizes from the Kaggle website. Using all these methods helped improve the accuracy.

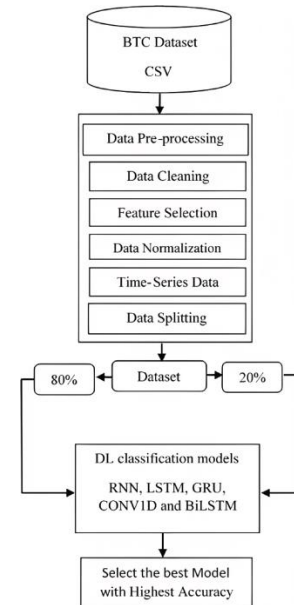


Fig. 6. The proposed model for predicting BTC price.

V. PROPOSED MODEL TESTING

During this research, an experiment was conducted to test five DL models, (RNN), (LSTM), (GRU), (Bi-LSTM), and (CONV1D); the models are designed to predict BTC Price.

A. Dataset

The data used in this study were downloaded from the Kaggle website for Bitcoin Cryptocurrency in CSV format. The dataset contains a variety of columns, including, Open, high, low, close, and Adj close prices and the volume, from the period 2014-09-17 to 2022-02-01, as shown in Fig. 7, the sample data from the datasets of the Cryptocurrency used in the study, and the target variable in this research is only the (Close Prices) Bitcoin.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2014-09-17	465.864014	468.174011	452.421997	457.334015	457.334015	21056800
1	2014-09-18	456.859985	456.859985	413.104004	424.440002	424.440002	34483200
2	2014-09-19	424.102997	427.834991	384.532013	394.795990	394.795990	37919700
3	2014-09-20	394.673004	423.295990	389.882996	408.903992	408.903992	36863600
4	2014-09-21	408.084991	412.425995	393.181000	398.821014	398.821014	26580100

Fig. 7. Historical prices for bitcoin before the preprocessing.

B. Data Pre-Processing

Data preparation is the first step of the experiment. Before data is supplied to the DL models, preprocessing is considered a crucial step that must be finished. Many stages were conducted during the processing, including data cleaning, feature selection, data normalization, time-series data, and data splitting.

- **Data Cleaning:** Replacing null or incorrect values with legitimate ones or eliminating the whole data point.
- **Feature Selection:** There are several variables in cryptocurrency data. Since only the Close Prices are the goal variable in our study, only relevant characteristics should be chosen, and extraneous features should be eliminated, as shown in Fig. 8.
- **Data normalization or standardization:** These processes are crucial for ensuring all data is on the same scale. In our models, we employ the MinMaxScaler function to normalize all data to a range of 0 to 1, which is crucial when working with Bitcoin data, which might have a broad range of values.
- **Time-series Data:** Cryptocurrency prices are time series data; as a result, it is crucial to transform the data into a time series format to recognize any patterns, trends, or seasonal impacts.
- **Data Splitting:** Divide the pre-processed data into training and testing sets. The prediction model will be developed using the training dataset, and its effectiveness will be assessed using the testing dataset.

	Date	Close
0	2021-02-06	39266.011719
1	2021-02-07	38903.441406
2	2021-02-08	46196.464844
3	2021-02-09	46481.105469
4	2021-02-10	44918.183594

Fig. 8. The target variable (Close Prices).

C. Deep Learning Models for Predicting BTC Price Movement

We propose in this section five DL algorithms. (1) RNN (2) LSTM (3) GRU (4) BiLSTM (5) CONV1D. The architecture of these models is shown in Tables II to VI.

1) **Recurrent Neural Network (RNN) model:** Simple RNN is the model's first layer; it has one simple RNN layer consisting of 50 filters that acquire data, process it, and then pass it on to the next layer. The results are in a 7 x 50 matrix using ReLU as an Activation Function. The second layer is another simple RNN, which generates a 1 x 20 matrix using ReLU as an Activation Function. Next, the last stage in the model consists of two fully connected layers, the first one with 50 nodes and the last one with one node, which is the model's output, and we used the Adam optimizer to calculate the learning rate, as shown in Table II.

TABLE II. RECURRENT NEURAL NETWORK (RNN) MODEL

Layer (type)	Output shape	Param #
simple_rnn (SimpleRNN)	(None, 7, 50)	2600
simple_rnn_1 (SimpleRNN)	(None, 20)	1420
dense_4 (Dense)	(None, 50)	1050
dense_5 (Dense)	(None, 1)	51
Total params : 5,121 Trainable params : 5,121 Non-trainable params : 0		

2) **Long Short-Term Memory (LSTM) model:** LSTM is the second DL model; LSTM is the model's first layer consisting of 50 filters that acquire data, process it, and then pass it on to the next layer. The results are in a 7 x 50 matrix using ReLU as an Activation Function. The second layer is another LSTM, which generates a 1 x 25 matrix using ReLU as an Activation Function. Next, the last stage in the model consists of two fully connected layers: the first one with 50 nodes and the last with one node, which is the model's output, and the Adam optimizer method, as shown in Table III.

TABLE III. LONG SHORT-TERM MEMORY (LSTM) MODEL

Layer (type)	Output shape	Param #
lstm_5 (LSTM)	(None, 7, 50)	10400
lstm_6 (LSTM)	(None, 25)	7600
dense_12 (Dense)	(None, 50)	1300
dense_13 (Dense)	(None, 1)	51
Total params : 19,351 Trainable params : 19,351 Non-trainable params : 0		

3) *Gate Recurrent Unit (GRU) model*: GRU is the third DL model, GRU is the model's first layer which generates a 1 x 50 matrix, and the last stage in the model is composed of two fully connected layers, the first one with 50 nodes and the last with one node which is the output of the model, and the Adam optimizer method as shown in Table IV.

TABLE IV. GATE RECURRENT UNIT (GRU) MODEL

Layer (type)	Output shape	Param #
gru (GRU)	(None, 50)	7950
dense_8 (Dense)	(None, 50)	2550
dense_9 (Dense)	(None, 1)	51
Total params : 10,551 Trainable params : 10,551 Non-trainable params : 0		

4) *Bidirectional Long Short-Term Memory (Bi-LSTM) Model*: Bi-LSTM is the fourth DL model. Bi-LSTM is the model's first layer consisting of 200 filters that acquire data, process it, and then pass it on to the next layer. The results are in a 207 x 200 matrix. A dropout layer is a regularization approach that prevents overfitting problems in deep learning by ensuring that no units are codependent with one another. Next, the last stage in the model is composed of two fully connected layers, the first using ReLU as an Activation Function with 20 nodes and the last with one node, which is the model's output, and the Adam optimizer method, as shown in Table V.

TABLE V. BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM) MODEL

Layer (type)	Output shape	Param #
bidirectional (Bidirectional)	(207, 200)	81600
dropout (Dropout)	(207, 200)	0
Dense_10 (Dense)	(207, 20)	4020
dense_11 (Dense)	(207, 1)	21
Total params : 85,641 Trainable params : 85,641 Non-trainable params : 0		

5) *1D Convolutional Neural Network (CONVID) model*: 1DCNN is the fifth DL model; 1DCNN is the model's first layer consisting of 64 filters that acquire data, process it, and then pass it on to the next layer. The results are in a 7 x 64 matrix which uses ReLU as an Activation Function. In order to simplify the output and avoid overfitting the data, the maximum pooling layer is used after a CNN layer; This indicates that the output matrix for this layer is 2 x 64 in size. The Max Pooling1D layer shrinks the input representation by taking the maximum value across all time dimensions. Next, the last stage in the model is composed of two fully connected layers, the first with 50 nodes, then using the Flatten layer; the Flatten layer transforms convolutional layer output into a single, one-dimensional vector that may be utilized as the input for a dense layer. The last dense layer has one node, the model's output, and the Adam optimizer method, as shown in Table VI.

TABLE VI. 1D CONVOLUTIONAL NEURAL NETWORK (CONVID) MODEL

Layer (type)	Output shape	Param #
conv1d (Conv1D)	(None, 7, 64)	512
max_pooling1d(Maxpooling1D)	(None, 2, 64)	0
dense_6 (Dense)	(None, 2, 50)	3250
flatten (Flatten)	(None, 100)	0
dense_7 (Dense)	(None, 1)	101
Total params : 3,863 Trainable params : 3,863 Non-trainable params : 0		

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The experiment's results will be discussed in this section.

A. Model Training

To find the best DL model, we trained utilizing DL models on the dataset in the first phase, splitting it into two groups of 80% training and 20% testing. Four assessment measures—RMSE, MSE, MAE, and R^2 —were used to examine and contrast the DL models, as will discuss in Section 6(C).

B. Epochs

The number of training set iterations is called an "epoch." The model's capacity for generalization improves as epochs increase. However, if the number of epochs is excessively high, an overfitting issue is readily created, and the model's capacity for generalization is diminished [42]. Therefore, picking the appropriate number of epochs is crucial. In this research, we used 200 epochs.

Tables VII, to XI show the loss and val_loss for each epoch on the various DL models. As shown in Fig. 9 to 13, the model's loss for the training and validation phases decrease in each epoch, indicating that the model performs optimally. The model predicts the actual and prediction phases shown in Fig. 14 to 18.

TABLE VII. LOSS, VAL LOSS OF RNN MODEL

Epoch	Loss	Val_Loss
1/200	0.0858	0.0042
2/200	0.0039	0.0032
3/200	0.0041	0.0033
4/200	0.0031	0.0037
5/200	0.0034	0.0034

TABLE VIII. LOSS, VAL LOSS OF LSTM MODEL

Epoch	Loss	Val_Loss
1/200	0.0654	0.0153
2/200	0.0110	0.0068
3/200	0.0082	0.0102
4/200	0.0082	0.0065
5/200	0.0073	0.0059

TABLE X. LOSS, VAL LOSS OF GRU MODEL

Epoch	Loss	Val_Loss
1/200	0.0214	0.0039
2/200	0.0040	0.0032
3/200	0.0039	0.0029
4/200	0.0038	0.0031
5/200	0.0035	0.0026

TABLE XI. LOSS, VAL LOSS OF BI-LSTM MODEL

Epoch	Loss	Val_Loss
1/200	0.0242	0.0082
2/200	0.0104	0.0095
3/200	0.0098	0.0089
4/200	0.0092	0.0049
5/200	0.0087	0.0125

TABLE XII. LOSS, VAL LOSS OF CONV1D MODEL

Epoch	Loss	Val_Loss
1/200	0.0160	0.0045
2/200	0.0049	0.0051
3/200	0.0046	0.0031
4/200	0.0037	0.0046
5/200	0.0044	0.0029

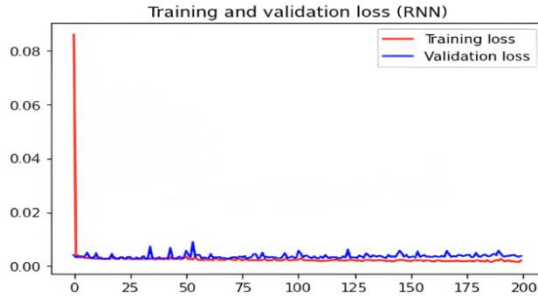


Fig. 9. RNN model loss for training and validation.

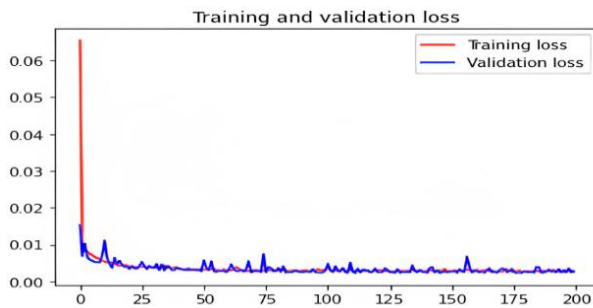


Fig. 10. LSTM model loss for training and validation.

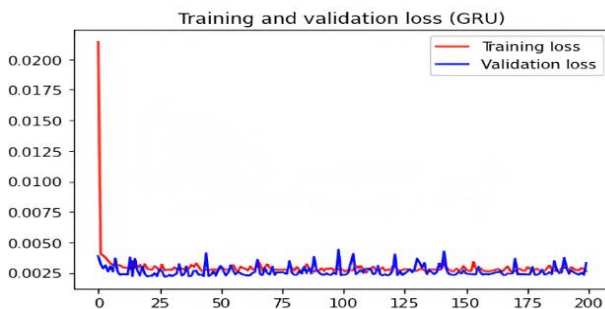


Fig. 11. GRU model loss for training and validation.

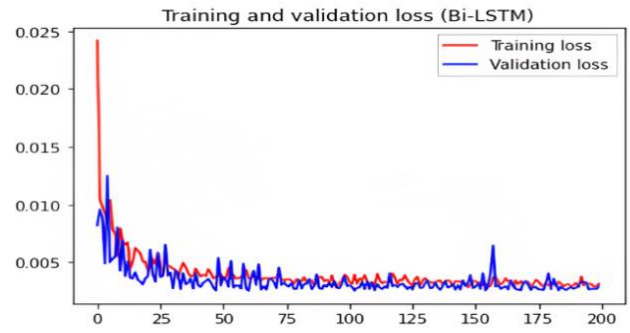


Fig. 12. Bi-LSTM model loss for training and validation.

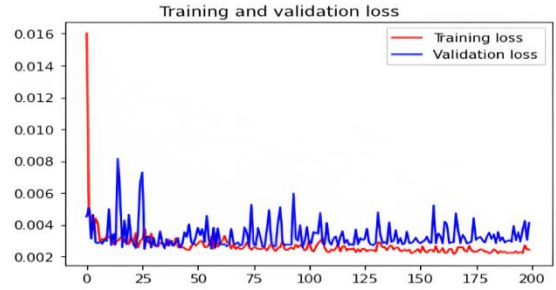


Fig. 13. CONV1D model loss for training and validation.

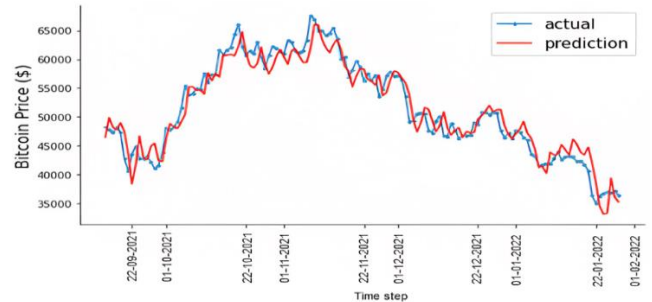


Fig. 14. BTC price prediction based on RNN model.

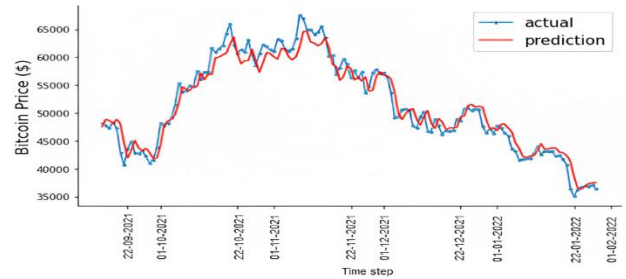


Fig. 15. BTC price prediction based on LSTM model.

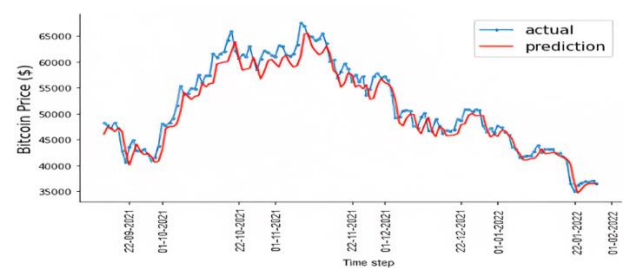


Fig. 16. BTC price prediction based on GRU model.

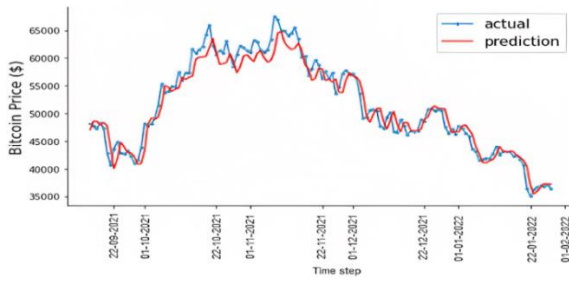


Fig. 17. BTC price prediction based on Bi-LSTM model.

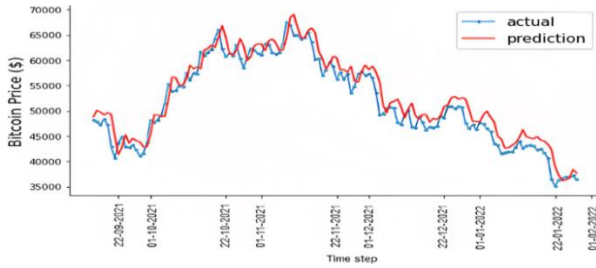


Fig. 18. BTC price prediction based on CONV1D model.

C. Evaluation Metrics

R-squared score (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) were used to assess the performance of the deep learning models.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_{t_i} - \hat{y}_{t_i})^2}{n}} \quad (1)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{t_i} - \hat{y}_{t_i})^2 \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{t_i} - \hat{y}_{t_i}| \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{t_i} - \hat{y}_{t_i}| \quad (4)$$

D. Deep Learning Prediction Models Outcomes

This section covered the outcomes of the prediction models created utilizing DL models between 22-09-2021 and 01-02-2022. How well prediction models perform Tables XII, to XVI, show the testing and training results regarding RMSE, MAE, MSE, and R^2 . Table XVII displays the outcomes of utilizing various models, with the model with the lowest error values chosen as the best model. The comparison between the actual and expected values for BTC price prediction models is shown in Fig. 19, as well.

TABLE XIII. TESTING AND TRAINING OUTCOMES FOR RNN MODEL

	RMSE	MAE	MSE	R^2
Training	1512.57653	1042.15566	2287887.78387	0.9739
Testing	2312.72885	1855.64295	5348714.75799	0.9232

TABLE XIV. TESTING AND TRAINING OUTCOMES FOR LSTM MODEL

	RMSE	MAE	MSE	R^2
Training	1908.46769	1476.50991	3642248.95956	0.95846
Testing	1978.68268	1537.14424	3915185.15068	0.94383

TABLE XV. TESTING AND TRAINING OUTCOMES FOR GRU MODEL

	RMSE	MAE	MSE	R^2
Training	2091.59478	1631.35273	4374768.76158	0.95011
Testing	2170.99032	1693.30095	4713198.99972	0.93238

TABLE XVI. TESTING AND TRAINING OUTCOMES FOR Bi-LSTM MODEL

	RMSE	MAE	MSE	R^2
Training	1882.75025	1462.76491	3544748.50956	0.95957
Testing	2048.97955	1574.88476	4198317.23545	0.93977

TABLE XVII. TESTING AND TRAINING OUTCOMES FOR CONV1D MODEL

	RMSE	MAE	MSE	R^2
Training	2039.41352	1535.66479	4159207.51736	0.95257
Testing	2418.95978	1949.64524	5851366.42758	0.91606

Table XVII shows that the LSTM model, which has the lowest RMSE, MAE, and MSE values and the greatest R^2 value, performs the best in forecasting BTC prices. Fig. 19, which demonstrate how closely the forecasts of the LSTM model match the actual prices, support this. The findings show that LSTM is a better predictor than RNN, GRU, Bi-LSTM, and CONV1D. The second and third-best models are the Bi-LSTM and GRU, with higher RMSE, MAE, and MSE values.

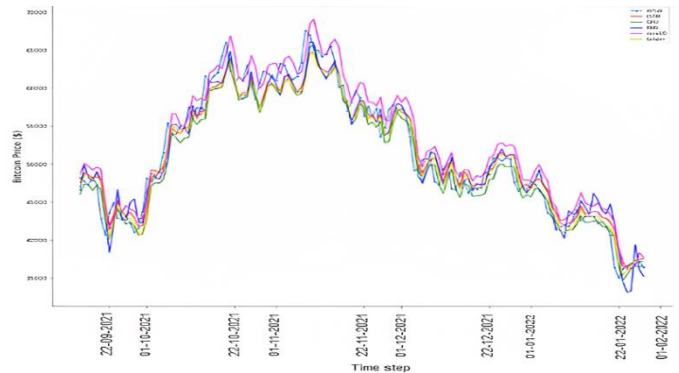


Fig. 19. Summary of BTC price prediction models between actual and predicting.

TABLE XVIII. SUMMARY OF DIFFERENT DL MODELS PREDICTION IN TERMS OF THE VARIOUS CRITERIA

Model	RMSE	MAE	MSE	R^2
RNN	2312.72885	1855.64295	5348714.75799	0.92327
Conv1D	2418.95978	1949.64524	5851366.42758	0.91606
GRU	2170.99032	1693.30095	4713198.99972	0.93238
Bi-STM	2048.97955	1574.88476	4198317.23545	0.93977
LSTM	1978.68268	1537.14424	3915185.15068	0.94383

These models are reliable and appropriate based models are reliable and appropriate based on the assessment techniques and outcomes. It should be emphasized that these models contain many flaws that may affect how well they can forecast BTC values:

- As cryptocurrency values rely heavily on various factors, LSTMs, RNN, GRU, Bi-LSTM, and CONV1D may only be able to account for some of these

dependencies, producing predictions that could be better.

- These models are vulnerable to overfitting, particularly when trained on small datasets, which can lead to subpar performance when used with new data.

VII. CONCLUSION AND FUTURE WORK

In this research, the market capitalization of the BTC cryptocurrency was utilized to forecast the price using five different deep learning techniques: LSTM, RNN, GRU, Bi-LSTM, and CONV1D. RMSE, MAE, MSE, and R2 values were used to assess the models' performance. The study's findings showed that the LSTM model, followed by the Bi-LSTM and GRU models, offered the best accurate forecasts for the price of the BTC coin. The study's findings show that deep learning algorithms are good at forecasting cryptocurrency values and that the LSTM model outperforms RNN, GRU, Bi-LSTM, and CONV1D.

To increase the precision of BTC predictions, the researcher plans to apply more deep learning algorithms or hybrid DL models in the future. The epoch size might also be increased to get a greater accuracy rate. Deep learning methods will also examine how emotion and tweets affect BTC pricing.

The research limitations can be represented in the following points:

- The prediction process focused on Bitcoin only. It did not apply the prediction to other cryptocurrencies, for example, Ethereum and Litecoin, which can correlate and impact the price of Bitcoin.
- Not considering another factor that can impact the rise and fall of a currency's price, such as comments on social media.

REFERENCES

- [1] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach," *Fractal and Fractional*, vol. 7, no. 2, pp. 203, 2023.
- [2] A. A. Oyedele, A. O. Ajayi, L. O. Oyedele, S. A. Bello, and K. O. Jimoh, "Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction," *Expert Systems with Applications*, vol. 213, pp. 119233, 2023.
- [3] N. Latif, J. D. Selvam, M. Kapse, V. Sharma, and V. Mahajan, "Comparative Performance of LSTM and ARIMA for the Short-Term Prediction of Bitcoin Prices," *Australasian Accounting, Business and Finance Journal*, vol. 17, no. 1, pp. 256-276, 2023.
- [4] S. A. Basher, and P. Sadorsky, "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?," *Machine Learning with Applications*, vol. 9, pp. 100355, 2022.
- [5] Z. Ye, Y. Wu, H. Chen, Y. Pan, and Q. Jiang, "A stacking ensemble deep learning model for bitcoin price prediction using Twitter comments on bitcoin," *Mathematics*, vol. 10, no. 8, pp. 1307, 2022.
- [6] N. Alsalmi, S. Ullah, and M. Rafique, "Accounting for digital currencies," *Research in International Business and Finance*, vol. 64, pp. 101897, 2023.
- [7] S. H. Hasan, S. H. Hasan, M. S. Ahmed, and S. H. Hasan, "A Novel Cryptocurrency Prediction Method Using Optimum CNN," *Computers, Materials & Continua*, vol. 71, no. 1, pp. 1051-1063, 2022.
- [8] E. Şaşmaz, and F. B. Tek, "Tweet sentiment analysis for cryptocurrencies," In 2021 6th International Conference on Computer Science and Engineering (UBMK), pp. 613-618, IEEE, 2021.
- [9] A. S. Salama, and A. M. Eassa, "IOT and cloud based blockchain model for COVID-19 infection spread control," *J Theor Appl Inf Technol*, 100(1), 113-126, 2022.
- [10] H. F. Nematallah, A. A. H. Sedky, and K. M. Mahar, "Bitcoin Price Trend Prediction Using Deep Neural Network," *Webology (ISSN: 1735-188X)* Vol. 19, 2022.
- [11] M. J. Hamayel, and A. Y. Owda, "A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms," *AI*, vol. 2, no. 4, pp. 477-496, 2021.
- [12] M. A. Ammer, and T. H. Aldhyani, "Deep learning algorithm to predict cryptocurrency fluctuation prices: Increasing investment awareness," *Electronics*, vol. 11, no. 15, pp. 2349, 2022.
- [13] B. Agarwal, P. Harjule, L. Chouhan, U. Saraswat, H. Airan, and P. Agarwal, "Prediction of dogecoin price using deep learning and social media trends," *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems*, vol. 8, no. 29, e2-e2, 2021.
- [14] E. Bouri, S. J. H. Shahzad, and D. Roubaud, "Co-explosivity in the cryptocurrency market," *Finance Research Letters*, 29, pp. 178-183, 2019.
- [15] S. Fakharchian, "Designing forecasting assistant of the Bitcoin price based on deep learning using the market sentiment analysis and multiple feature extraction," 2022.
- [16] F. Fang, C. Ventre, M. Basios, L. Kanthan, D. Martinez-Rego, F. Wu, and L. Li, "Cryptocurrency trading: a comprehensive survey," *Financial Innovation*, vol. 8, no. 1, pp. 1-59, 2022.
- [17] S. McNally, J. Roche, and S. Caton, "Predicting the price of bitcoin using machine learning," In 2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP), pp. 339-343, IEEE, 2018.
- [18] K. Murray, A. Rossi, D. Carraro, and A. Visentin, "On Forecasting Cryptocurrency Prices: A Comparison of Machine Learning, Deep Learning, and Ensembles," *Forecasting*, vol. 5, no. 1, pp. 196-209, 2023.
- [19] X. Jiang, "Bitcoin price prediction based on deep learning methods," *Journal of Mathematical Finance*, vol. 10, no. 1, pp. 132-139, 2019.
- [20] E. Mahendra, H. Madan, S. Gupta, and S. V. Singh, "Bitcoin price prediction using deep learning and real time deployment," In 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), pp. 264-268, IEEE, 2020.
- [21] T. Awoke, M. Rout, L. Mohanty, and S. C. Satapathy, "Bitcoin price prediction and analysis using deep learning models," In *Communication Software and Networks: Proceedings of INDIA 2019* (pp. 631-640). Singapore: Springer Singapore, 2020.
- [22] A. K. Bitto, I. Mahmud, M. H. I. Bijoy, F. T. Jannat, M. S. Arman, M. M. H. Shohug, and H. Jahan, "CryptoAR: scrutinizing the trend and market of cryptocurrency using machine learning approach on time series data," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 3, pp. 1684-1696, 2022.
- [23] S. Zhang, M. Li, and C. Yan, "The Empirical Analysis of Bitcoin Price Prediction Based on Deep Learning Integration Method," *Computational Intelligence and Neuroscience*, 2022, 2022.
- [24] I. Gurrib, and F. Kamalov, "Predicting bitcoin price movements using sentiment analysis: a machine learning approach," *Studies in Economics and Finance*, vol. 39, no. 3, pp. 347-364, 2022.
- [25] S. Cavalli, and M. Amoretti, "CNN-based multivariate data analysis for bitcoin trend prediction," *Applied Soft Computing*, 101, pp. 107065, 2021.
- [26] M. Liu, G. Li, J. Li, X. Zhu, and Y. Yao, "Forecasting the price of Bitcoin using deep learning," *Finance research letters*, vol. 40, pp. 101755, 2021.
- [27] S. Marne, S. Churi, D. Correia, and J. Gomes, "Predicting Price of Cryptocurrency-A deep learning approach," *NTASU-9* (3), 2020.

- [28] R. Karlemstrand, and E. Leckström, "Using Twitter attribute information to predict stock prices,". arXiv preprint arXiv:2105.01402, 2021.
- [29] B. A. Chandio, A. S. Imran, M. Bakhtyar, S. M. Daudpota, and J. Baber, "Attention-based RU-BiLSTM sentiment analysis model for Roman Urdu,". Applied Sciences, vol. 12, no. 7, pp. 3641, 2022.
- [30] K. Irie, R. Csordás, and J. Schmidhuber, "The dual form of neural networks revisited: Connecting test time predictions to training patterns via spotlights of attention,". In International Conference on Machine Learning (pp. 9639-9659). PMLR, 2022.
- [31] A. P. Rodrigues, R. Fernandes, A. Shetty, K. Lakshmana, and R. M. Shafi, "Real-time twitter spam detection and sentiment analysis using machine learning and deep learning techniques,". Computational Intelligence and Neuroscience, 2022, 2022.
- [32] N. R. Bhowmik, M. Arifuzzaman, and M. R. H. Mondal, "Sentiment analysis on Bangla text using extended lexicon dictionary and deep learning algorithms,". Array, vol. 13, pp. 100123, 2022.
- [33] B. Aditya Pai, L. Devareddy, S. Hegde, and B. S. Ramya, "A time series cryptocurrency price prediction using lstm,". In Emerging Research in Computing, Information, Communication and Applications: ERCICA 2020, Vol. 2, pp. 653-662. Springer Singapore, 2022.
- [34] D. Liu, and A. Wei, "Regulated LSTM Artificial Neural Networks for Option Risks,". FinTech, vol. 1, no. 2, pp. 180-190, 2022.
- [35] M. Bilgili, N. Arslan, A. ŞEKERTEKİN, and A. YAŞAR, "Application of long short-term memory (LSTM) neural network based on deeplearning for electricity energy consumption forecasting,". Turkish Journal of Electrical Engineering and Computer Sciences, vol. 30, no. 1, pp. 140-157, 2022.
- [36] C. Bai, "AGA-GRU: An optimized GRU neural network model based on adaptive genetic algorithm,". In Journal of Physics: Conference Series (Vol. 1651, No. 1, p. 012146). IOP Publishing, 2020.
- [37] K. Zarzycki, and M. Ławryńczuk, "Advanced predictive control for GRU and LSTM networks,". Information Sciences, vol. 616, pp. 229-254, 2022.
- [38] G. Xu, W. Guo, and Y. Wang, "Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-Conv architecture,". Medical & Biological Engineering & Computing, vol. 61, no. 1, pp. 61-73, 2023.
- [39] H. Elfaik, and E. H. Nfaoui, "Deep bidirectional LSTM network learning-based sentiment analysis for Arabic text,". Journal of Intelligent Systems, vol. 30, no. 1, pp. 395-412, 2020.
- [40] K. Yousaf, and T. Nawaz, "A deep learning-based approach for inappropriate content detection and classification of youtube videos,". IEEE Access, vol. 10, pp. 16283-16298, 2022.
- [41] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier,". Journal of Signal Processing Systems, vol. 91, pp. 179-189, 2019.
- [42] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM,". Ieee Access, vol. 7, pp. 51522-51532, 2019.

Advances in Value-based, Policy-based, and Deep Learning-based Reinforcement Learning

Haewon Byeon

Department of Medical Big Data-College of AI Convergence,
Inje University, Gimhae 50834, Gyeongsangnamdo, South Korea

Abstract—Machine learning is a branch of artificial intelligence in which computers use data to teach themselves and improve their problem-solving abilities. In this case, learning is the process by which computers use data and algorithms to build models that improve performance, and it can be divided into supervised learning, unsupervised learning, and reinforcement learning. Among them, reinforcement learning is a learning method in which AI interacts with the environment and finds the optimal strategy through actions, and it means that AI takes certain actions and learns based on the feedback it receives from the environment. In other words, reinforcement learning is a learning algorithm that allows AI to learn by itself and determine the optimal action for the situation by learning to find patterns hidden in a large amount of data collected through trial and error. In this study, we introduce the main reinforcement learning algorithms: value-based algorithms, policy gradient-based reinforcement learning, reinforcement learning with intrinsic rewards, and deep learning-based reinforcement learning. Reinforcement learning is a technology that enables AI to develop its own problem-solving capabilities, and it has recently gained attention among AI learning methods as the usefulness of the algorithms in various industries has become more widely known. In recent years, reinforcement learning has made rapid progress and achieved remarkable results in a variety of fields. Based on these achievements, reinforcement learning has the potential to positively transform human lives. In the future, more advanced forms of reinforcement learning with enhanced interaction with the environment need to be developed.

Keywords—Reinforcement learning; value-based algorithms; policy gradient-based reinforcement learning; reinforcement learning with intrinsic rewards; deep learning-based reinforcement learning

I. INTRODUCTION

Advances in artificial intelligence and machine learning technologies have led to the development and use of AI-based services in many industries. At the same time, models using reinforcement learning, a branch of machine learning, are growing rapidly.

Machine learning is a branch of artificial intelligence in which computers use data to teach themselves and improve their problem-solving abilities. In this case, learning is the process by which computers use data and algorithms to build models that improve performance, and it can be divided into supervised learning, unsupervised learning, and reinforcement learning [1, 2]. Among them, reinforcement learning [3] is a learning method in which AI interacts with the environment and finds the optimal strategy through actions, and it means

that AI takes certain actions and learns based on the feedback it receives from the environment [4]. In other words, reinforcement learning is a learning algorithm that allows AI to learn by itself and determine the optimal action for the situation by learning to find patterns hidden in a large amount of data collected through trial and error. Reinforcement learning is a technology that enables AI to develop its own problem-solving capabilities, and it has recently gained attention among AI learning methods as the usefulness of the algorithms in various industries has become more widely known [5, 6, 7].

This study is structured as follows. Section II presents the history and components of reinforcement learning, and Section III describes the main reinforcement learning algorithms: Value Based Algorithms, Policy Gradient Based Reinforcement Learning, Reinforcement Learning with Intrinsic Reward, and Reinforcement Learning based on Deep Learning. Section IV describes applications of reinforcement learning. Finally, Section V presents trends in the application of Reinforcement Learning in networking and future research directions. Section VI outlines the limitations and Section VII presents the conclusion to the study.

II. HISTORY AND COMPONENTS OF REINFORCEMENT LEARNING

Reinforcement learning can be traced back to an optimisation method for solving sequential decision problems, mathematically modelled by the Markov Decision Process, developed in the 1950s [8]. A Markov decision process is defined as a tuple (S, A, P, R, γ) , where S and A are the agent's state space and action space, respectively. P and R are transition probability and reward functions, respectively, where P is the probability distribution of the next state and R is the reward the agent will receive in the next state if it performs an action $a \in A$ in state $S \in S$. The reward is a metric for judging the goodness or badness of the agent's actions. γ is the discount rate used to write off future rewards when calculating cumulative rewards. This helps the agent reach the goal quickly and prevents the cumulative reward from drifting, so that learning is stable [9]. The optimal policy for the sequential decision problem defined by the Markov Decision Process presented above can be found by reinforcement learning.

A policy is a function that takes a state value as input and determines what action the agent should take. A reinforcement learning agent observes the state of itself and its environment based on its sensors and information from other agents, decides what to do based on the observed state values and policies, and is sometimes rewarded for its actions. At this point, the

reinforcement learning agent learns its policy to maximise its cumulative reward expectation, and the policy that maximises the cumulative reward expectation is the optimal policy. This is based on the Reward Hypothesis. The reward hypothesis states that any goal (e.g., solving a problem) can be described as maximising the agent's cumulative reward.

This implies the importance of designing a reward function. Since reinforcement learning is how an agent learns its policy by interacting with the environment through trial and error, relying on reward cues, a poor reward function will not only make it difficult for the agent to learn, but will also lead to unexpected side effects even if it does learn [10].

In summary, reinforcement learning generally consists of three main elements.

- Agent: An object that performs actions in the environment according to the current policy.
- Environment: The external system with which the agent interacts. The environment provides feedback to the agent in the form of rewards or punishments based on actions taken.
- Reward: Feedback is the positive or negative feedback an agent receives from the environment based on its actions.

The agent's goal is to maximise the total amount of rewards over time, which means that the goal of reinforcement learning is to find a strategy that maximises the cumulative reward in a given environment [11]. In most cases, this means emphasising long-term rewards over short-term rewards. This process of reinforcement learning can be thought of as a trial-and-error process, where the AI learns by taking actions and observing the rewards or punishments that result from those actions. The agent uses this information to update its policies so that it can make better decisions in the future.

III. MAIN REINFORCEMENT LEARNING ALGORITHMS

Reinforcement learning algorithms can be classified into value-based, policy-based, and model-based algorithms.

A. Value-based Algorithms

Value-based algorithms estimate the value of each state or state-action pair and select the optimal action to improve the agent's performance [12]. Typical examples are Q-learning and Deep Q-Network (DQN)(Fig. 1 and Fig. 2).

Q-learning based reinforcement learning approximates the Q-value for a state-action pair each time and then decides which action to take in which state [13]. For exploration, we often use ϵ -greedy policies. An ϵ -greedy policy is a method that chooses a random action in a given state with probability ϵ , and the action with the highest Q-value with probability $(1 - \epsilon)$. Representative examples are DQN and Rainbow [14], which combines DQN with six DQN improvement algorithms.

The technical features of DQN can be summarised as follows: first, the use of convolutional neural networks for image recognition; second, the introduction of empirical replay to eliminate the correlation between samples and increase the efficiency of sampling; and third, the separation of the online

Q-network, which determines the agent's behaviour, and the target Q-network, which is used to calculate the target Q-value, for learning stability [15].

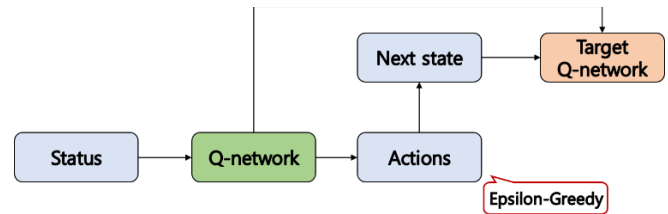


Fig. 1. The concept of deep Q network.

First, Rainbow is a technique that combines DQN with the following six DQN enhancement algorithms. For example, double Q-learning DQN takes the maximum value of the target Q-network in the current state when calculating the target Q-value, resulting in overestimation of the target Q-value and poor learning performance. Double Q-learning prevents the overestimation of the target Q-value by calculating the target Q-value using the behavioural value that maximises the online Q-net as the input of the target Q-net [16].

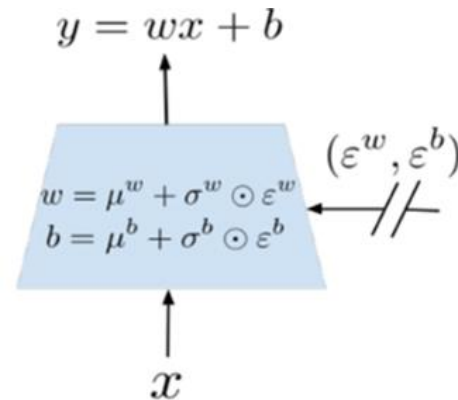


Fig. 2. Concept of applying gaussian noise to the weights of a neural network.

Second, prioritised experience replay - DQN learns by extracting experiences uniformly from experience replay, i.e. prioritised experience replay is a method of extracting samples that are more likely to be conducive to learning [17].

Third, Dueling Networking-DQN calculates Q-values on the fly. Since the Q-value takes into account both state and behaviour, it can be strongly influenced by behaviour when evaluating the value in a particular state. By decomposing the Q-value into the value of a particular state and the benefit of different actions that can be taken in that state, Dueling Networks can compute the value of a particular state more robustly while taking into account the value of actions [18].

Fourth, multi-step learning DQN uses the reward after the 1-step bootstrap, which is the very next state, to compute the target Q-value. By extending it to learn with the reward information after n-step bootstrapping, it evolves into multi-step learning (ex. with improved learning stability and speed) [19].

Fifth, Distributional RL-DQN, uses the expectation of the Q-value. In this case, the limitation is that it is difficult to

exploit the randomness inherent in the Markov decision process if only the expectation of the Q-value is used. In this case, Distributional RL is a method that uses a distribution of rewards instead of a single average. By using Distributional RL, you can not only improve learning performance, but also design safer agents by allowing agents to avoid risky behaviours [20].

Sixth, Noisy Nets-DQN uses the ϵ -greedy policy. However, the ϵ -greedy policy often leads to inefficient exploration because it outputs random behaviour regardless of the agent's current situation, and there is also the problem of setting the ϵ value. This is where Noisy Nets can be used (Fig. 3). By training a neural network (Noisy Nets) that adds noise to the weights and biases of the policy neural network when training the policy neural network, Noisy Nets has the advantage that the randomness of the agent's behaviour automatically adapts to the state the agent is in and over time (reducing randomness and promoting greedy choices as training progresses) [21].

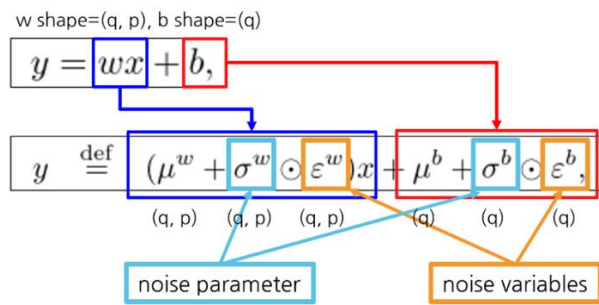


Fig. 3. The concept of noisy networks.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \odot \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11} b_{11} & a_{12} b_{12} & a_{13} b_{13} \\ a_{21} b_{21} & a_{22} b_{22} & a_{23} b_{23} \\ a_{31} b_{31} & a_{32} b_{32} & a_{33} b_{33} \end{bmatrix}$$

Fig. 4. The concept of element-wise multiplication.

B. Policy Gradient-based Reinforcement Learning

A policy-based algorithm directly optimises the policy that determines the agent's behaviour. In other words, policy gradient reinforcement learning directly yields a policy that determines which action to take in which state. A policy can be defined as a parameter vector θ , which is a function that takes observations as input and outputs an action value as output. The use of probabilistic policies in reinforcement learning is efficient for balancing exploration and exploitation. In deep reinforcement learning, this is approximated using a functional rate deep neural network, where the parameter vector θ is the weight and bias of the neural network. Policy gradient based reinforcement learning uses a policy gradient technique to compute this parameter vector, θ . The policy gradient technique is a method that uses gradient multiplication to find θ (Fig. 4). In other words, after finding the gradient of the objective function for a given θ , updating θ by a certain distance in the direction of the increasing gradient is repeated until the gradient converges or for a maximum time step.

The objective function is the expectation of the cumulative reward of acting according to the policy, expressed as in Eq. (1).

$$J(\theta) = E_{\pi_{\theta}}[r(s, a)] \quad (1)$$

The gradient of the objective function is defined by the policy gradient theorem [22], as shown in Eq. (2).

$$\nabla J(\theta) = E_{\pi_{\theta}}[Q_{\pi}(s, a) \nabla_{\theta} \ln \pi_{\theta}(a|s)] \quad (2)$$

The policy gradient updates the policy's parameters in the direction of maximising the objective function, as shown in Eq. (3).

$$\theta \leftarrow \theta + a \nabla J(\theta) \quad (3)$$

A prime example of policy-based reinforcement learning is Proximal Policy Optimisation (PPO) [22]. One of the drawbacks of policy-based reinforcement learning is that the policy of the parameters can change rapidly. This leads to learning instability, which results in slow learning speed and poor performance. To prevent such learning instability, TRPO (Trust region policy optimisation) [23] adds a condition that constrains the Kullback-Leibler (KL) divergence before and after a policy update to be below a certain level. TRPO attracted the attention of researchers due to its success in solving robot control problems with continuous action spaces, which were not solved by DQN. However, TRPO has the disadvantage that it requires a lot of computation to solve the constraints and is incompatible with various neural network structures (e.g., dropout, parameter sharing). To compensate for these disadvantages while maintaining the performance of TRPO, PPO removes the computationally intensive KL divergence constraint and indirectly limits the number of policy renewals by simply clipping the ratio of pre- and post-policy renewals in the TRPO objective function. Although PPO was published in 2017, it is still the state-of-the-art algorithm and has been reported in many studies to be very good in terms of performance, computational efficiency and ease of implementation [24].

C. Reinforcement Learning with Intrinsic Reward

Reinforcement learning is a method of learning by exploring the environment through trial and error. In this case, the agent evaluates its behaviour and updates its policy based on the rewards it receives as a result of its actions. This can work well in environments where every action is rewarded, but policy learning is less successful in environments where rewards are infrequent and darkness is delayed. For example, suppose a game has a single reward for avoiding all obstacles, skeletons, etc. to get the key. In this case, it is difficult to determine whether an action taken in a particular state to get the key was beneficial, harmful or pointless. Even DQNs that outperform humans in many games are likely to fail in the game described above. One way to deal with this problem is to use intrinsic rewards.

Intrinsic rewards are rewards that are generated by the agent rather than given by the environment. It mimics the way humans learn through intrinsic motivation. A typical intrinsic reward function is prediction error. Prediction error is defined as the difference between the agent's predicted next state and

the actual next state. To predict the next state, an agent typically defines one or more prediction models (e.g., artificial neural networks) and trains them along with a policy. As the agent explores the environment and learns the prediction models, this prediction error will be lower for states that are familiar to the agent and higher for states that are unfamiliar to the agent, which has the effect of encouraging the agent to explore and, in games, discouraging the agent from dying and returning to the initial state. This is because the initial state is familiar to the agent, as it is the state to which the agent returns when it dies. The intrinsic reward function is designed so that the agent receives a reward for every action it performs, so it can learn well even in environments where the environment is sparse and black is a delayed reward. This also helps to broaden the application of reinforcement learning, as it reduces the need for a human to design a precise reward function for each task in the environment. However, there are a number of issues that need to be considered, such as the match between the directionality of the task the agent needs to perform and the directionality of the internal reward, the non-stationarity of the reward for performing the same behaviour in the same state as learning progresses, and the scaling of the reward across different environments.

To illustrate this, we refer the reader to Random Network Distillation [25]. Random Network Distillation consists of three neural networks: goal, prediction, and policy. The policy neural network is the one that determines the agent's behaviour, while the goal and prediction neural networks take the next state value as input and output some feature value. The goal neural network is fixed with randomly set weights, and the prediction neural network is a neural network with the same structure as the goal neural network, and is trained together with the policy neural network to produce the same output as the goal neural network. In other words, it is called Random Network Distillation because it has the effect of distilling a random network into a predictive neural network. In Random Network Distillation, the internal reward value function and the external reward value function are obtained separately and then combined, and PPO is used to optimise the policy neural network.

D. Reinforcement Learning (RL) Based on Deep Learning

Reinforcement learning refers to a group of methods for solving stochastic decision problems. Reinforcement learning can be classified as a model belonging to supervised learning or as an independent field of reinforcement learning. The reason it is classified as supervised learning is that it receives feedback or guidance from the environment, including humans, as it learns. On the other hand, it is classified as an independent model because the optimal decision process of reinforcement learning is a learning model that differs from the label-based discriminative approach typical of supervised learning. Reinforcement learning is very similar to the way humans learn, as it uses a trial and error process. For this reason, the core algorithm of AlphaGo, developed by Google DeepMind, is based on reinforcement learning. This makes reinforcement learning the closest model to artificial intelligence. Unlike supervised learning, reinforcement learning is not given training data. Instead, the reinforcement learning problem is given a reward function. The definition of solving

reinforcement learning is to find a policy function that maximises the average of future reward values. To solve reinforcement learning, researchers have borrowed a mathematical model: the Markov decision process (MDP). Intuitively, the Markov property means that, given the present, the past and the future are independent. For example, the score I get on a test tomorrow depends only on my current state and how much I study today. An MDP has four main parts.

- A set of states
- A set of actions
- A transition function
- A reward function

In this study, we will illustrate the above four points with the example in Fig. 5. The most common example used to describe MDPs is the situation of a robot in a lattice space as shown below.

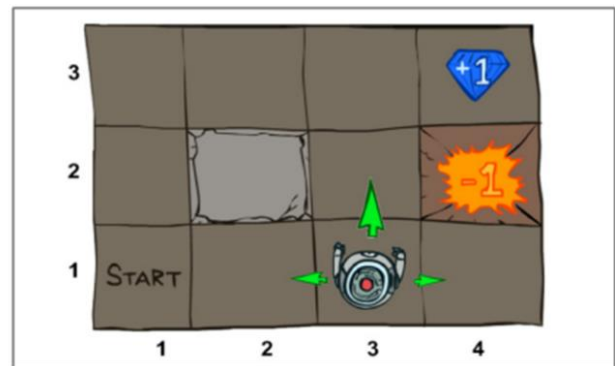


Fig. 5. A robot in a lattice space.

As shown in Fig. 5, the robot can be in one of twelve grids, which is its state space. In each grid the robot can move up, down, left, right or stay in place, and these five actions correspond to the action space. If the robot's movement through the grid is determined, it will move in the desired direction, but if it is stochastic, even if the robot tries to move up, it will move right or left with some probability. In this way, the transition function describes the probability of reaching the next state when a particular action is performed in a particular state. Finally, and most importantly for reinforcement learning, a reward function is defined for each state: in the above lattice space, reaching a gem is rewarded with a +1 reward, and reaching a fire is rewarded with a -1 reward. Given an MDP, solving reinforcement learning means finding a policy function that maximises the sum of expected future rewards. This has several implications, but the most important is that it involves future rewards. While rewards from current behaviour are important, ultimately we need to consider both current and future rewards. The next thing to remember is that we're dealing with a stochastic system. It is possible that our current behaviour will not lead us to the desired state, which means that if we can get the same reward, it is better to get it 'sooner'. The discount factor takes this into account. It is set to a value less than 1 and the reward earned over time is multiplied by this value.

So, given an MDP, how do we find the optimal policy function? The two most basic ways of solving reinforcement learning are value iteration and policy iteration. To explain this, we first need to define value. If we know not only the immediate reward we can get now, but also the consensus expectation of the rewards we can get when we start from that state, we can choose the action that maximises that function each time, and thus find the optimal policy function. This consensus expectation of future rewards is called the value function, $V(s)$. Similarly, given a current state the expected future reward for taking an action in the current state is called the action value function or Q-function, $Q(s,a)$. Value iteration refers to the method of finding this value function. The value function is difficult to define intuitively because it is not only about the current state, but also about future states and the rewards that can be obtained in those states. In general, reinforcement learning uses the Bellman equation to find this value function, which is defined as follows [26].

$$V(s) = \max_a \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V(s')] \quad (4)$$

We can see that the above formula is a recursive equation with the value function $V(s)$ we want to find on the left and right sides, and the rest is the same as MDP except for $V(s)$. We can initialise $V(s)$ to any value and run the above recursive equation for all states s until it converges, always finding the optimal $V(s)$. In the previous section we looked at value iteration, but now we want to look at policy iteration, which is different from value iteration. Policy iteration consists of two phases: policy evaluation, which evaluates the performance of the current policy function, and policy improvement, which improves the policy based on the evaluation. These two phases alternate until the policy function converges. It is generally accepted that policy iterations converge to the optimal policy function faster than value iterations.

In this section we will investigate how to find the optimal policy function in an MDP when no model is given. This problem is commonly referred to as model-free reinforcement learning. The main difference from the model-based reinforcement learning described earlier is that we no longer know how the environment behaves. In other words, you do something in one state and get a reward for the next state, which is "passively" informed by the environment. Model-free reinforcement learning has several differences from model-based reinforcement learning, the most important of which is exploration. Since we don't know how the environment will behave, we have to experiment and use the results to gradually learn the policy function. Let's see how we can solve model-free reinforcement learning defined in this way. We can't use the Bellman equation directly because we don't know $T(s, a, s')$, which is the part of the Bellman equation used in model-based reinforcement learning.

Policy evaluation is a methodology for evaluating a given $V(s)$, replacing a with $\pi(s)$ above.

$$V(s) = \max_a \sum_{s'} T(s, \pi(s), s') [R(s, \pi(s), s') + \gamma V(s')] \quad (5)$$

The above formula is the policy evaluation formula used in model-free reinforcement learning. $T(s, \cdot (s), s')$ is an unknown value, but if the next state, s' , comes from a model called T , we can replace the sum of T with the sample mean. One method

that replaces the Bellman equation with sampling in this way is temporal difference (TD) learning. In an MDP, experience means that in a state s , I take an action via a given policy function ($a = \pi(s)$), and as a result I receive the next state s' and a reward r . This reward r is a function of (s, a, s') . As the experiences of (s, a, s', r) accumulate, we learn a value function $V(s)$ and an action value function $Q(s,a)$ based on these data. The expression for the Bellman equation for $Q(s,a)$ is as follows.

$$Q(s, a) \leftarrow \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma \max_{a'} Q(s', a')] \quad (6)$$

The above formula can be used to update the behavioural value function $Q(s,a)$. In this formula, we can replace T with the sample estimate, which gives us the following formula.

$$(1 - \alpha)Q(s, a) + \alpha [R(s, a, s') + \gamma \max_{a'} Q(s', a')] \quad (7)$$

Solving reinforcement learning problems using the above formula is called Q-learning. All that is needed to update the Q-function using this formula is the experience of (s, a, s', r) , i.e. the action taken in one state, the observation of the next state, and the reward received, and the Q-function can be obtained using the above formula. The advantage of Q-learning is that it can find the optimal policy function without knowing the model. However, Q-learning is not a one-size-fits-all method. The main disadvantage of model-less reinforcement learning is exploration. Due to the nature of reinforcement learning, rewards may be given only once or sporadically at the end. The difficulty of this exploration increases as the state space grows.

In deep reinforcement learning (DRL) [27], a method that combines this Q-learning with deep learning techniques, one of the most important things is how well it explores. One such method is called Deep Q Network (DQN), made famous by DeepMind. DQN is the algorithm that enabled DeepMind's AlphaGo to win the 2016 World Championship against Lee Sedol and Kasparov. From the algorithmic point of view of traditional reinforcement learning, DQN doesn't bring much new to the table. However, the main advantage of the QL formula is that it can update the Q function without any information about the environment. This property is more important than you might think, because the difficulty of RL compared to traditional SL is that it considers the sum of expected future rewards, but the above formula allows you to update the Q-function with just one experience. It also has the advantage that when choosing (s, a) from (s, a, s', r) , $Q(s, a)$ always converges to the optimal action value function after infinite time, even if a random action a is chosen each time. The right-hand side of the above formula is easy to find if we know the current $Q(s, a)$ function. To find $\max_{a'} Q(s', a')$, the number of possible actions a must be finite. In other words, we can plug in all the possible actions and pick the one with the largest Q value. And if we think of $Q(s, a)$ on the left side as the output of the Q function for some input (s, a) and the right side as the target for that input, the Bellman equation for the Q function above can be interpreted as giving us the input-output pairs for the regression function that we often use in supervised learning.

IV. APPLICATIONS OF REINFORCEMENT LEARNING

Recently, there has been a lot of research and development on reinforcement learning, one of the artificial intelligence algorithms, to solve network system optimisation problems. Reinforcement learning is a system control method in which a reinforcement learning agent in a network management system uses information derived from the network environment to construct a reward function and achieve an optimal goal through iterative improvement. To do this, reinforcement learning agents go through an organic process of changing the state of the environment, controlling the behaviour of the agent, designing the value function, designing the reward function, improving the policy, and deriving the optimisation model. However, in order to learn a value function for decision making by predicting the expected value of output through predefined states and actions, a large amount of time must be invested in learning, and learning may not be performed well due to excessive environmental state information provided, or learning may be performed with the wrong goal. To overcome these problems, reinforcement learning models that improve learning efficiency and prediction accuracy performance by configuring the system's reward function as an artificial intelligence neural network have been studied. In addition, reinforcement learning models that can perform effective learning not only for discrete and limited number of behaviours, but also for very high degrees of freedom of the behaviours to be controlled are being studied.

V. TRENDS IN THE APPLICATION OF REINFORCEMENT LEARNING IN NETWORKING

As the network structure becomes more complex, various problems in the areas of routing, resource management, security and QoS/QoE arise, and to solve them, reinforcement learning application techniques are being studied, which include adaptive optimisation mechanisms for different environments. In the area of routing, research is being carried out to use reinforcement learning to optimise the routing process as network traffic grows exponentially. In resource management, researchers are applying reinforcement learning for efficient resource management and scheduling in rapidly changing network environments such as smart cities or edge clouds. This enables efficient network management by controlling network congestion or reducing overhead. In the area of network security, reinforcement learning is used to detect and respond to anomalies in the network, such as network congestion. In the area of QoS/QoE, researchers are using reinforcement learning to improve overall QoS/QoE by taking into account dynamically changing network characteristics.

VI. LIMITATIONS OF REINFORCEMENT LEARNING

One of the limitations of reinforcement learning is low data efficiency, especially in tasks where data selection is expensive, time consuming or dangerous. Therefore, one of the ways to deal with this is the off-policy technique, which can overcome the limitations of reinforcement learning to some extent if the behaviour policy and the target policy are different and the behaviour policy is carefully learned. Under this premise, imitation learning can also achieve good performance. However, most imitation learning algorithms have difficulties

in achieving performance in suboptimal trajectory situations, and usually require interaction with the environment to overcome them. Therefore, in the future, it is necessary to develop imitation learning with enhanced interaction with the environment that can overcome suboptimal trajectory situations.

VII. CONCLUSION

In summary, reinforcement learning is a learning method at the heart of the AI revolution, enabling unimagined innovations in fields as diverse as autonomous driving, healthcare and gaming. As with any ML technology, it is important to consider the ethical implications of its use and ensure that it is applied in a responsible and beneficial way. In recent years, reinforcement learning has made rapid progress and achieved remarkable results in a variety of fields. Based on these achievements, reinforcement learning has the potential to positively transform human lives. In the future, more advanced forms of reinforcement learning with enhanced interaction with the environment need to be developed.

ACKNOWLEDGMENT

This research Supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- RS-2023-00237287, NRF-2021S1A5A8062526) and local government-university cooperation-based regional innovation projects (2021RIS-003).

REFERENCES

- [1] M. I. Jordan, T. M. Mitchell, Machine learning: Trends, perspectives, and prospects. *Science*, vol. 349, no. 6245, pp. 255-260, 2015.
- [2] B. Mahesh, Machine learning algorithms-a review. *Int. j. sci. res.*, vol. 9, no. 1, pp. 381-386, 2020.
- [3] Y. Li, Deep reinforcement learning: An overview. *arXiv preprint*, 2017.
- [4] K. Arulkumar, M. P. Deisenroth, M. Brundage, A. A. Bharath, Deep reinforcement learning: A brief survey. *IEEE Signal Process Mag.*, vol. 34, no. 6, pp. 26-38, 2017.
- [5] R. Nian, J. Liu, B. Huang, A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*, vol. 139, pp. 106886, 2020.
- [6] P. Dayan, Y. Niv, Reinforcement learning: the good, the bad and the ugly. *Curr. Opin. Neurol.*, vol. 18, no. 2, pp. 185-196, 2008.
- [7] M. Botvinick, S. Ritter, J. X. Wang, Z. Kurth-Nelson, C. Blundell, D. Hassabis, Reinforcement learning, fast and slow. *Trends Cogn. Sci.*, vol. 23, no. 5, pp. 408-422, 2019.
- [8] M. L. Puterman, Markov decision processes. *Handbooks in operations research and management science*, vol. 2, pp. 331-434, 1990.
- [9] W. S. Lovejoy, A survey of algorithmic methods for partially observed Markov decision processes. *Ann. Oper. Res.*, vol. 28, no. 1, pp. 47-65, 1991.
- [10] C. Guestrin, M. Lagoudakis, R. Parr, Coordinated reinforcement learning. In *ICML*, Vol. 2, pp. 227-234, 2002.
- [11] J. Oh, M. Hessel, W. M. Czarnecki, Z. Xu, H. P. van Hasselt, S. Singh, D. Silver, Discovering reinforcement learning algorithms. In *Neural Information Processing Systems*, vol. 33, pp. 1060-1070, 2020.
- [12] X. Zang, H. Yao, G. Zheng, N. Xu, K. Xu, Z. Li, Metalight: Value-based meta-reinforcement learning for traffic signal control. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 01, pp. 1153-1160, 2020.
- [13] A. Kumar, A. Zhou, G. Tucker, S. Levine, Conservative q-learning for offline reinforcement learning. In *Neural Information Processing Systems*, vol. 33, pp. 1179-1191, 2020.

- [14] M. Hessel, J. Modayil, H. Van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, et al. Rainbow: Combining improvements in deep reinforcement learning. In Proceedings of the AAAI conference on artificial intelligence. Vol. 32, No. 1, pp. 3215-3222, 2018.
- [15] Y. Wang, H. Liu, W. Zheng, Y. Xia, Y. Li, P. Chen, et al. Multi-objective workflow scheduling with deep-Q-network-based multi-agent reinforcement learning. IEEE access, vol. 7, pp. 39974-39982, 2019.
- [16] K. Arulkumaran, M. P. Deisenroth, M. Brundage, A. A. Bharath, Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, 2017.
- [17] X. Wang, H. Xiang, Y. Cheng, Q. Yu, Prioritised experience replay based on sample optimisation. J. Eng., vol. 13, pp. 298-302, 2020.
- [18] N. Van Huynh, D. T. Hoang, D. N. Nguyen, E. Dutkiewicz, Optimal and fast real-time resource slicing with deep dueling neural networks. IEEE J. Sel. Areas Commun., vol. 37, no. 6, pp. 1455-1470, 2019.
- [19] J. F. Hernandez-Garcia, R. S. Sutton, Understanding multi-step deep reinforcement learning: A systematic study of the DQN target. arXiv preprint, 2019.
- [20] Y. Tang, R. Munos, M. Rowland, B. Avila Pires, W. Dabney, M. Bellemare, The nature of temporal difference errors in multi-step distributional reinforcement learning. In Neural Information Processing Systems, vol. 35, pp. 30265-30276, 2022.
- [21] S. Han, W. Zhou, J. Liu, S. Lü, NROWAN-DQN: A stable noisy network with noise reduction and online weight adjustment for exploration. arXiv preprint, 2020.
- [22] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. arXiv preprint, 2017.
- [23] J. Schulman, S. Levine, P. Abbeel, M. Jordan, P. Moritz, Trust region policy optimization. In International conference on machine learning, pp. 1889-1897, 2015.
- [24] W. Yi, R. Qu, L. Jiao, Automated algorithm design using proximal policy optimisation with identified features. Expert Syst. Appl., vol. 216, pp. 119461, 2023.
- [25] Y. Burda, H. Edwards, A. Storkey, O. Klimov, Exploration by random network distillation. arXiv preprint, 2018.
- [26] B. O'Donoghue, I. Osband, R. Munos, V. Mnih, The uncertainty bellman equation and exploration. In International Conference on Machine Learning, pp. 3836-3845, 2018.
- [27] T. Zahavy, M. Haroush, N. Merlis, D. J. Mankowitz, S. Mannor, Learn what not to learn: Action elimination with deep reinforcement learning. In neural information processing systems, vol. 31, 2018

Scalable Blockchain Architecture: Leveraging Hybrid Shard Generation and Data Partitioning

Praveen M Dhulavvago¹, Prasad M R², Niranjana C Kundur³, Jagadisha N⁴, S G Totad⁵

School of Computer Science and Engineering, KLE Technological University, Hubli, India¹⁻⁵

Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru, India²

Department of Computer Science and Engineering, JSS Academy of Technical Education, Bengaluru, India³

Department of Information Science and Engineering, Canara Engineering College Bantwal, India⁴

Abstract—Blockchain technology has gained widespread recognition and adoption in various domains, but its implementation beyond crypto currencies faces a significant challenge - poor scalability. The serial execution of transactions in existing blockchain systems hampers transaction throughput and increases network latency, limiting overall system performance. In response to this limitation, this paper proposes a static analysis-driven data partitioning approach to enhance blockchain system scalability. By enabling parallel and distributed transaction execution through a simultaneous block-level transaction approach, the proposed technique substantially improves transaction throughput and reduces network latency. The study employs a hybrid shard generation algorithm within the Geth node of the blockchain network to create multiple shards or partitions. Experimental results indicate promising outcomes, with miners experiencing a remarkable speedup of 1.91x and validators achieving 1.90x, along with a substantial 35.34% reduction in network latency. These findings provide valuable insights and offer scalable solutions, empowering researchers and practitioners to address scalability concerns and promoting broader adoption of blockchain technology across various industries.

Keywords—Ethereum; shard generation; data partitioning; proof of work

I. INTRODUCTION

Blockchain technology has emerged as a groundbreaking innovation with transformative potential across various industries, revolutionizing the way we conduct transactions and manage data. The decentralized and tamper-resistant nature of blockchain networks, first introduced with Bit coin in 2008 by the enigmatic Satoshi Nakamoto, has paved the way for new applications beyond crypto currencies, such as supply chain management, healthcare, finance, and more. However, as blockchain networks continue to gain traction and see widespread adoption, a critical challenge looms large: scalability. Scalability is a pivotal concern in the broader implementation of blockchain systems [1]. As the number of participants (network nodes) and the volume of transactions grow exponentially, traditional blockchain architectures face limitations in accommodating the increasing demands on their resources and processing capabilities. This has led to performance bottlenecks, increased network latency, and limited transaction throughput, hindering the seamless scalability of blockchain networks. To address this challenge, this paper presents a novel and comprehensive solution that

leverages a combination of hybrid shard generation and data partitioning techniques. The primary objective of this approach is to enhance the scalability of blockchain architectures while preserving the core principles of decentralization, security, transparency, and immutability [2].

Sharding techniques and data partitioning have emerged as promising solutions to address scalability challenges in blockchain networks. By effectively distributing the workload and data across multiple shards or partitions, these techniques aim to enhance transaction throughput and overall network performance [3]. However, implementing sharding and data partitioning in blockchain networks comes with its own set of challenges. Ensuring consensus across multiple shards, facilitating secure cross-shard communication, and maintaining data synchronization are critical considerations.

The proposed solution begins with the introduction of a groundbreaking hybrid shard generation algorithm, inspired by the concept of sharding in database technology. This algorithm intelligently creates multiple shards within the blockchain network, each functioning as an independent blockchain with its own set of verified users and data. By strategically distributing the transaction load among these shards, we aim to optimize resource utilization and improve overall system performance, facilitating seamless scalability. Moreover, to further enhance the efficiency of the blockchain network, we employ a static analysis-driven data partitioning technique. This innovative approach enables the execution of transactions in parallel and distributed fashion across different shards, reducing contention and enhancing coordination among network nodes. By effectively partitioning data and workloads, we mitigate performance bottlenecks, promoting smoother transaction processing and improved scalability [4].

The combination of hybrid shard generation and data partitioning forms a powerful and synergistic approach to tackle the scalability challenge in blockchain networks. By enhancing transaction throughput, reducing network latency, and efficiently utilizing network resources, our solution aims to pave the way for the broader adoption of blockchain technology across diverse industries. The primary goal of this study is to implement the sharding technique in a blockchain network to improve transaction throughput and reduce network latency. By distributing the workload across multiple shards, the aim is to achieve higher transaction processing capacity and faster transaction confirmation times.

The two significant contributions to enhance the scalability of blockchain networks are:

- A novel hybrid shard generation algorithm is introduced, creating multiple shards within the network, each functioning as an independent blockchain with its set of verified users and data. The algorithm strategically distributes transaction load among these shards, optimizing resource utilization and improving system performance.
- A static analysis-driven data partitioning technique is proposed, enabling parallel transaction execution across shards, reducing contention and enhancing efficiency.

The paper is organized with discussions on related work in Section II, followed by the detailed design of the hybrid shard generation algorithm in Section III, explanation of the static analysis-based data partitioning technique in Section IV, and an experimental evaluation in Section V, showcasing impressive speedups for miners and validators and reduced network latency. A comparison with existing solutions highlights the advantages and effectiveness of the proposed approach. Section VI concludes the study.

II. RELATED WORK

This section provides a comprehensive review of the existing literature and research on blockchain scalability, shedding light on the current state of the field and the various approaches proposed to tackle scalability challenges. The related work encompasses studies on sharding techniques, data partitioning, and other scalability solutions tailored for blockchain networks. Satoshi Nakamoto's introduction of blockchain in 2008 [5] numerous industries such as finance, healthcare, supply chain, and real estate have experienced significant benefits from its innovative capabilities. However, both traditional distributed databases and blockchain systems exhibit distinct failure modes, as highlighted in [6]. To address this, a shard formation mechanism has been developed to facilitate the creation of efficient shards for parallel processing. The process of mining, being laborious and computationally intensive, has led to the emergence of distributed mining pools in well-known blockchain systems like Bitcoin and Ethereum. These pools harness distributed computational power to collectively identify the Proof of Work (PoW) and distribute rewards based on pre-established protocols [7]. Notably, the concept of dynamic blockchain sharding has been proposed, enabling the blockchain to alter its sharding configuration in real-time without requiring hard forks. An exemplary implementation of dynamic blockchain sharding is found in RapidChain.

Yizhong Liu et al. [8] present a thorough investigation of sharding in blockchain systems, with a focus on fundamental concepts, diverse approaches, and the challenges related to blockchain scalability. Their study offers a comprehensive overview of the essential building blocks of sharding solutions, with a particular emphasis on partitioning strategies, consensus mechanisms, and data management. Hung Dang et al. [9] have made significant progress in extending sharding to permissioned blockchain systems, emphasizing its potential as

a solution to address scalability challenges in blockchain networks. Their primary goal is to develop a blockchain system capable of accommodating large network size comparable to major crypto currencies like Bitcoin and Ethereum. By doing so, they aim to overcome the limitations of existing blockchain networks, which are often confined to crypto currency applications and struggle to scale consensus protocols for handling average workloads comparable to centralized processing systems[10].

Mahdi Zamani et al. [11] introduce Rapid Chain as a robust and Byzantine-resilient public blockchain protocol. By employing full sharding, Rapid Chain efficiently partitions network nodes into multiple committees, enabling parallel processing of disjoint blocks of transactions and maintaining separate ledgers. This sharding approach brings significant advancements in scalability for blockchain systems. The authors further conduct a comprehensive performance comparison, pitting Rapid Chain against other state-of-the-art sharding-based protocols like Elastico and Omniledger [12]. The evaluation showcases RapidChain's smooth scalability, effectively supporting network sizes of up to 4000 nodes. Deepal Tennakoon et al. [13] introduce a novel dynamic blockchain sharding protocol, which offers advanced capabilities such as creating new shards, adjusting existing shards, and rotating shard participants. This dynamic approach addresses the limitations of traditional blockchain sharding protocols that lacked the flexibility to modify the number of shards used. The authors also emphasize the importance of security during shard creation to prevent malicious nodes from taking control of shards [14][16]. To counter bribery attempts, the paper proposes a shard committee rotation approach through transaction mapping. The performance evaluation of the solution on the CollaChain blockchain demonstrates quasi-linear scalability, highlighting its effectiveness in handling an increasing number of shards [18][19]. A comprehensive and systematic study of sharding techniques in blockchain systems. They extensively examine the key components of sharding schemes and the major challenges associated with each component. The paper thoroughly discusses various methods to generate epoch randomness and techniques for handling cross-sharding transactions [21]. This study provides valuable insights into the state-of-the-art in sharding on blockchain, contributing to a deeper understanding of this important scalability solution [15][17].

III. DESIGN OF HYBRID SHARD GENERATION ALGORITHM

The design of hybrid shard generation algorithm is a novel approach aimed at creating efficient and scalable blockchain networks through the intelligent generation of multiple shards. Hybrid shard generation algorithm is derived through a combination of concepts from traditional sharding techniques and data partitioning strategies, with the aim of optimizing blockchain network scalability. The algorithm follows a systematic process to intelligently create multiple shards within the network, ensuring efficient transaction processing and resource utilization. It is derived through a combination of concepts from traditional sharding techniques and data partitioning strategies, with the aim of optimizing blockchain network scalability [20]. Here is a detailed description of the key steps involved in the hybrid shard generation algorithm:

A. Shard Size Determination

The first step in the algorithm is to determine the ideal size of each shard. This is based on various factors, such as network capacity, computational power, and desired transaction throughput. The goal is to find a shard size that allows for efficient processing of transactions within each shard without causing performance bottlenecks.

B. High-Volume Transaction Identification

The technique analyzes the transaction data to identify high-volume transactions and frequently accessed smart contracts. These high-volume transactions are crucial for shard creation, as they form the basis for creating initial shards.

C. Shard Creation

Shard creation is formed using the identified high-volume transactions and relevant smart contracts, the technique forms initial shards. Each initial shard is designed to handle specific types of transactions efficiently. The technique assigns a unique identifier to each shard for easy reference.

D. Load Balancing

Once the shards are created the load balancing technique analyzes the transaction load on each shard. The goal is to evenly distribute the workload among the shards to optimize resource usage and avoid overloading any specific shard. If there is an imbalance in the transaction distribution, load balancing mechanisms are employed to address it.

E. Dynamic Sharding

One of the key features of the Hybrid Shard Generation Technique is its dynamic sharding capability. This means that the technique can adjust the number of shards in real-time based on changing network conditions and transaction demands. This adaptability allows the blockchain network to scale efficiently as the transaction load fluctuates.

F. Consensus Mechanism Selection

Once the shards are formed the consensus mechanism is applied on each shard, this mechanism defines an appropriate consensus mechanism that suits the specific requirements and characteristics of the transactions processed within that shard. Different shards may use different consensus protocols, depending on their unique needs.

G. Data Partitioning

Data partitioning strategies are implemented to ensure that relevant data is stored within each shard. The aim is to minimize the need for frequent cross-shard communication, as this can impact the overall performance of the blockchain network. Data is partitioned in a way that reduces data access across different shards.

H. Communication Protocol Establishment

Blockchain transactions may require interactions between different shards, the technique establishes a communication protocol to facilitate cross-shard transactions when necessary. This communication protocol ensures secure and efficient communication between the shards.

I. Security Measures

To protect the security and integrity of the blockchain network, the technique incorporates robust security measures. These measures are designed to prevent shard takeovers by malicious nodes and safeguard the overall security and decentralization of the network.

The hybrid shard generation algorithm continuously optimizes the shard configuration, load balancing, and consensus mechanisms based on the dynamic nature of the network. It adapts to changing transaction demands, ensuring that the blockchain network can efficiently scale while maintaining security and performance [16]. The result of the algorithm is a set of optimized shards that work collaboratively to achieve enhanced scalability, increased transaction throughput, and reduced network latency. Hybrid Shard Generation Algorithm is a technique used to intelligently create multiple shards within a blockchain network, optimizing transaction distribution and load balancing to enhance scalability.

The Parameters considered for the design of hybrid shard generation algorithm are as follows [24]:

- **Total Number of Shards:** This parameter defines how many shards will be created in the blockchain network. The number of shards affects the overall network capacity and scalability.
- **Shard Size:** Each shard's size determines the number of transactions it can accommodate. Smaller shard sizes might lead to more efficient processing, but could also introduce overhead due to the increased number of shards.
- **Hybrid Approach Ratio:** If the algorithm combines different shard generation approaches (e.g., static and dynamic), this parameter might define the proportion of each approach to use.
- **Dynamic Thresholds:** If dynamic shard generation is used, parameters related to the thresholds for triggering the creation or merging of shards might be defined.
- **Data Partitioning Criteria:** Parameters related to how transactions are partitioned into different shards, such as transaction attributes, geographical location, or other relevant factors.
- **Security and Consensus Parameters:** Depending on the consensus mechanism used in the blockchain network (e.g., Proof of Work, Proof of Stake), there might be parameters related to security, validation, and consensus that impact shard generation.
- **Load Balancing Criteria:** Parameters related to load balancing across shards to ensure even distribution of transactions and computational resources.
- **Adaptability Parameters:** Parameters that determine how the system adapts to changing conditions, such as variations in transaction volume or network size.

A high-level description of the Hybrid Shard Generation Algorithm and its pseudocode:

Algorithm: Hybrid shard generation algorithm

Input:

Total number of nodes in the blockchain network (N)
Desired number of shards (S)
Sharding criteria and parameters (e.g., transaction load, user verification)

Output:

List of shards with their assigned nodes

Procedure:

- a. Calculate the number of nodes per shard ($\text{Nodes_Per_Shard} = N / S$).
- b. Initialize an empty list to store shards and their assigned nodes (*Shard_Assignment*).
- c. For each shard ($i = 1$ to S), perform the following steps:
 - i. Create a new shard (*Shard_i*).
 - ii. Select *Nodes_Per_Share* nodes randomly from the blockchain network and assign them to *Shard_i*.
 - iii. Add *Shard_i* with its assigned nodes to the *Shard_Assignment* list.
- d. If there are any remaining nodes after all shards have been created:
 - i. Assign the remaining nodes to existing shards to balance the load (e.g., round-robin fashion).
 - ii. Update the *Shard_Assignment* list accordingly.
- e. Return the *Shard_Assignment* list as the final result.

```
def hybrid_shard_generation(N, S):  
    # Calculate the number of nodes per shard  
    nodes_per_shard = N // S  
    # Initialize an empty list to store shards and their assigned  
    # nodes  
    Shard_Assignment = [ ]  
    # Create shards and assign nodes  
    for i in range(S):  
        shard_i = "Shard_" + str(i+1)  
        assigned_nodes=randomly_select_nodes(N,nodes_per_shard)  
        shard_assignment.append ((shard_i, assigned_nodes))  
    # Handle any remaining nodes  
    remaining_nodes = N % S  
    for i in range(remaining_nodes):  
        Shard_Assignment[i][1].append(Nodes[i])  
    return Shard_Assignment
```

IV. STATIC ANALYSIS-BASED DATA PARTITIONING

Static Analysis-Based Data Partitioning is an innovative technique designed to enhance the efficiency and scalability of blockchain networks by optimizing data distribution and transaction execution across different shards. Unlike traditional data partitioning methods that rely on runtime analysis, this approach utilizes static analysis to pre-determine data partitioning strategies, enabling more effective workload distribution and minimizing contention among network nodes. Here is a detailed description of the Static Analysis-Based Data Partitioning technique

A. Data Analysis

The first step in this technique involves a comprehensive analysis of the blockchain data. The goal is to identify data patterns, dependencies, and access frequency for various transactions and smart contracts. Static analysis tools are employed to analyze the code and data structures within the blockchain network.

B. Dependency Graph Generation

Based on the data analysis, a dependency graph is generated. This graph represents the relationships between different data elements and their dependencies on each other. The dependency graph provides insights into how data should be logically grouped and distributed across different shards.

C. Workload Estimation

This technique estimates the workload for each shard based on the transactions and smart contracts that are likely to be executed within each shard. This workload estimation helps in ensuring that each shard is appropriately sized to handle its share of the transactions.

D. Transaction Grouping

Transaction grouping technique groups related transactions together based on their dependencies and access patterns. Transactions that frequently interact with the same data elements are grouped together to reduce the need for cross-shard communication.

E. Shard Allocation

Once transaction grouping operation is performed, the data is partitioned and allocated to specific shards based on the dependencies and workload estimation. The goal is to minimize data access across different shards, thereby reducing contention and enhancing transaction processing speed.

F. Parallel Transaction Execution

Once the data is partitioned and allocated to the shards, parallel transaction execution process is carried out within each shard. Transactions within the same shard can be executed concurrently, optimizing resource utilization and reducing transaction processing time.

G. Cross-Shard Communication Optimization

Although the focus is on minimizing cross-shard communication, some transactions may still require interactions with data in other shards. To optimize such communication, the technique employs efficient communication protocols and algorithms, reducing latency and ensuring secure and timely cross-shard interactions.

H. Scalability Evaluation

The performance of the blockchain network with Static Analysis-Based Data Partitioning is evaluated through experiments and simulations. The goal is to assess the scalability improvements achieved by this technique and compare it with other data partitioning approaches.

Static Analysis-Based Data Partitioning is a technique that optimizes the distribution of data and workloads across different shards in a blockchain network. The goal is to

improve the efficiency of transaction processing and reduce contention among network nodes. Static Analysis-based Data Partitioning offers a proactive and efficient approach to data distribution and transaction execution in blockchain networks. By leveraging static analysis tools and pre-determined partitioning strategies, it minimizes performance bottlenecks, reduces contention, and enhances overall system performance, contributing to the seamless scalability and improved efficiency of the blockchain network [25].

Here is the Pseudocode and algorithm for Static Analysis-Based Data Partitioning:

Algorithm: Static Analysis-Based Data Partitioning

Input:

Transaction pool: List of pending transactions in the blockchain network.

Shards: List of shards in the blockchain network.

Static analysis data: Information about the workload and data distribution of each shard.

Output:

Partitioned transactions: Transactions distributed across different shards based on static analysis.

Procedure:

- a. Initialize an empty dictionary to store the workload estimation for each shard.
- b. For each shard in the Shards list, perform the following steps:
 - i. Calculate the estimated workload for the shard using static analysis data, such as the number of pending transactions and their complexity.
 - ii. Store the estimated workload in the dictionary.
- c. Sort the shards in ascending order based on their estimated workload.
- d. For each transaction in the Transaction pool, perform the following steps:
 - i. Assign the transaction to the shard with the lowest estimated workload.
 - ii. Update the estimated workload for that shard in the dictionary.
- e. Return the partitioned transactions, which are now distributed across different shards based on static analysis.

```
# Initialize an empty dictionary to store the workload estimation for each shard  
workload_estimation = { }
```

```
# Calculate the estimated workload for each shard using static analysis data for shard in Shards:  
workload = calculate_workload_estimate(shard)  
workload_estimation[shard] = workload
```

```
# Sort the shards in ascending order based on their estimated workload  
sorted_shards=sort_shards_by_workload  
(workload_estimation)
```

```
# Initialize an empty dictionary to store the partitioned transactions  
partitioned_transactions = { }  
# Assign transactions to shards based on workload for transaction in Transaction_pool:  
shard_with_lowest_workload = sorted_shards[ ]  
partitioned_transactions[transaction]=shard_with_lowest_workload  
workload_estimation[shard_with_lowest_workload] += 1  
sorted_shards=sort_shards_by_workload(workload_estimation)  
# Re-sort the shards after workload update  
# Return the partitioned transactions  
return partitioned_transactions
```

V. RESULTS AND ANALYSIS

The experiment is carried out using historical transaction data from the Ethereum blockchain, obtained from Google's Bigquery Engine's public-data archive. It's important to note that the proposed hybrid shard generation algorithm might indeed be better suited to some types of data. Blockchain networks often deal with diverse use cases, such as financial transactions, supply chain data, IoT data, and more. A hybrid approach might be optimized for specific types of transactions or use cases, making it perform better in scenarios that align with its design goals. The dataset comprised approximately 5 million transactions distributed across 80,000 blocks. Within the dataset, two types of transactions were considered: monetary transactions, involving simple and low-latency transfers of value, and contractual transactions, which required the execution of smart contracts and took longer to process. Hyperledger Caliper is a benchmarking tool specifically designed to measure the performance of blockchain systems, including transaction scalability. It supports various blockchain platforms, including Ethereum and Hyperledger Fabric, which makes it suitable for evaluating the proposed technique in different blockchain environments.

To evaluate the proposed scalable blockchain architecture, a workload was created by varying the ratios of smart contract and monetary transactions within each block. Different ratios, such as {1/1, 1/2, 1/4, 1/8, 1/16}, were used to represent different transaction scenarios. Each block in the workload contained 140 to 560 transactions based on the specified ratios. The execution setups involved both serial execution and parallel execution with multiple slave setups.

A. Performance Evaluation Metrics

The key metrics considered for assessing the effectiveness of the proposed Hybrid Shard Generation Algorithm and Static Analysis-Based Data Partitioning technique are:

- **Transaction Throughput:** This metric quantified the number of transactions processed per second by the blockchain network. A higher transaction throughput indicates better efficiency in handling a larger volume of transactions within a given time frame. Improving transaction throughput is a crucial aspect of enhancing the scalability of blockchain networks [22].

- Network Latency: Network latency refers to the time taken for a transaction to be propagated and confirmed across the blockchain network. It was measured as the end-to-end block generation time, which indicates how quickly new blocks are created and validated. Lower network latency signifies faster confirmation and validation of transactions, contributing to a more responsive and efficient blockchain system [23].

These performance metrics are fundamental in evaluating the proposed technique's impact on the scalability of the blockchain network. By analyzing transaction throughput and network latency, The aim is to demonstrate improvements in transaction processing speed and reduced network latency, leading to a more scalable and responsive blockchain architecture.

Fig. 1 show the average transaction execution time by the miner in relation to the number of shards generated, ranging from 1 to 5. As the number of shards increases, the transaction execution time decreases, measured in milliseconds (ms). This reduction in transaction execution time is attributed to the increased transaction processing efficiency within a block, resulting in enhanced transaction throughput. A larger number of shards in a block enable higher levels of parallelism, facilitating faster transaction processing and contributing to improved overall system performance.

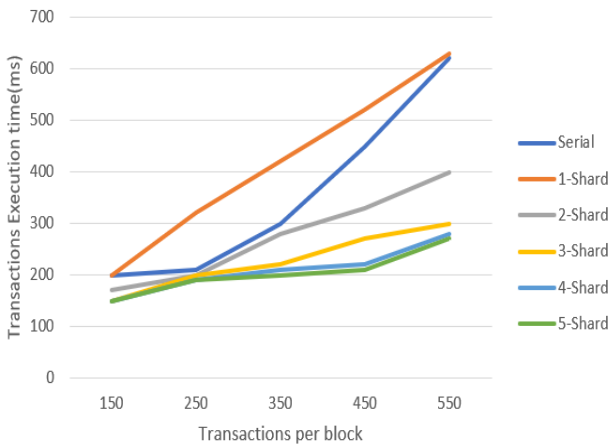


Fig. 1. Average transaction execution time by miner.

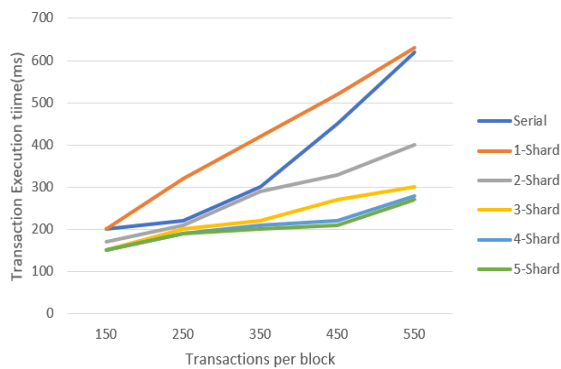


Fig. 2. Average transaction execution time by validator.

Fig. 2 illustrates the average transaction execution time by the validator as the number of shards generated increases from 1 to 5. The transaction execution time, measured in milliseconds (ms), decreases with a higher number of shards. This decrease in transaction execution time is attributed to the increased transaction processing efficiency within a block, resulting in improved transaction throughput. As the number of shards in a block increases, the transaction processing of the transactions within that block becomes more efficient, enabling greater parallelism. This enhancement in parallel processing contributes to the overall improvement in transaction throughput.

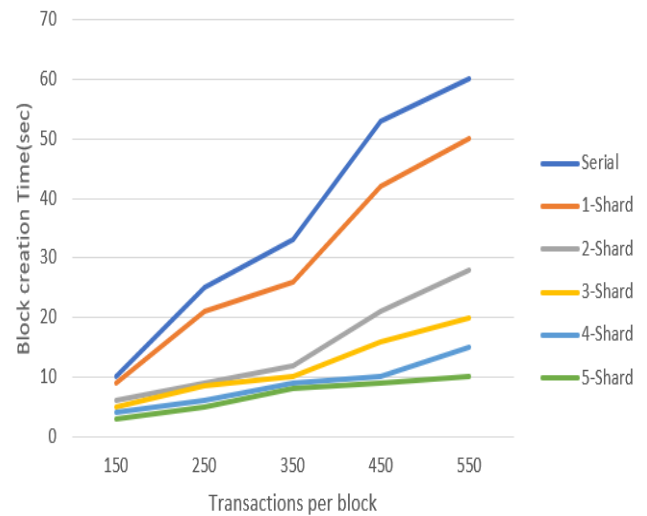


Fig. 3. Average end-to-end block creation time by miner (with mining).

Fig. 3 displays the average end-to-end block creation time by the miner (with mining) as the number of shards increases. As the number of shards rises, the block creation time, which refers to the time taken for a block to be accepted and added to the ledger, decreases. This reduction in block creation time is attributed to the efficient processing of more transactions at a faster rate when more shards are utilized. Consequently, the adoption of multiple shards leads to shorter block creation times, as the transactions are processed more rapidly and added to the blockchain in a more efficient manner.

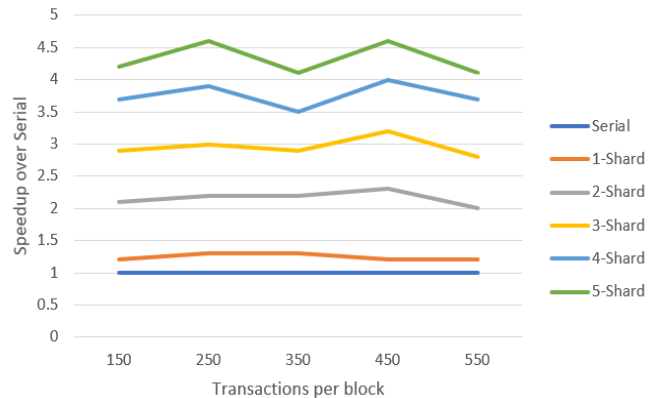


Fig. 4. Speedup of parallel mining over serial mining for average block creation.

Fig. 4 illustrates the speedup of parallel mining over serial mining for average block creation. Speedup represents the increase in transaction throughput compared to the baseline of serial execution, which is assigned a value of 1x. As the number of shards increase, the speedup observed becomes more significant, indicating a higher enhancement in transaction throughput. This increase in speedup is attributed to the efficient parallel processing of transactions achieved with the greater number of shards. Consequently, the adoption of more shards leads to a substantial boost in transaction throughput, demonstrating the effectiveness of the proposed approach in improving the overall performance of the blockchain system.

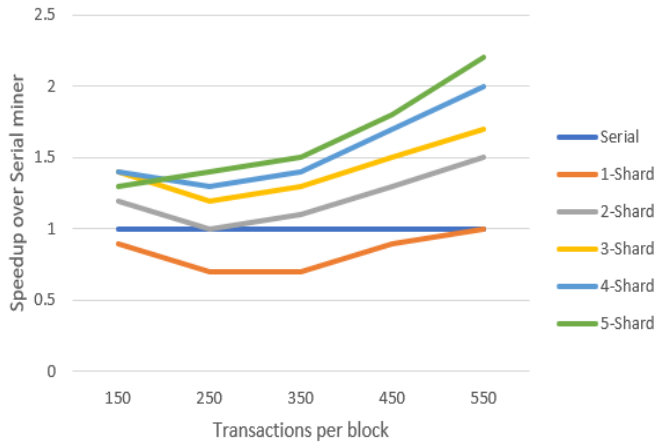


Fig. 5. Miner average speed increase (without mining) for transaction execution.

Fig. 5 displays the miner's average speed increase (without mining) for transaction execution. Speedup represents the rise in transaction throughput when compared to the baseline of serial execution, where the serial miner execution is assigned a value of 1x. As the number of shards increases, the observed speedup becomes more significant, indicating a higher enhancement in transaction throughput. Notably, the 5-shard architecture offers the highest speedup over the serial miner, suggesting that the adoption of a 5-shard configuration provides the most substantial improvement in transaction processing efficiency. This finding underscores the effectiveness of the proposed technique in optimizing the performance of the blockchain system and achieving higher transaction throughput.

Fig. 6 illustrates the validator's average speed increase for transaction execution. Speedup represents the rise in transaction throughput when compared to the baseline of serial execution, where the serial validator execution is assigned a value of 1x. As the number of shards increases, the observed speedup becomes more significant, indicating a higher enhancement in transaction throughput. Remarkably, the 5-shard architecture offers the highest speedup over the serial validator, signifying that the adoption of a 5-shard configuration provides the most substantial improvement in transaction processing efficiency for validators. This result further reinforces the efficacy of the proposed technique in optimizing the performance of the blockchain system and achieving higher transaction throughput for validators.

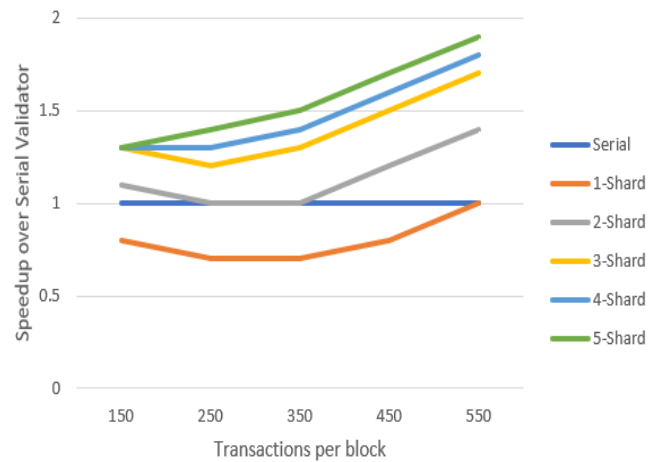


Fig. 6. Validator average speed increase for transaction execution.

B. Performance Analysis and Discussion

Table I presents the estimated workload speedup achieved with validators and miners (without mining) for different numbers of slaves in the system. The speedup values indicate how much faster the workload is executed when both validators and miners work concurrently compared to sequential execution (baseline). The table also provides the total number of transactions executed per block for each workload scenario. In the baseline scenario of serial execution, both miners and validators achieve a speedup of 1.00, which means no improvement in transaction throughput compared to sequential execution.

TABLE I. ESTIMATED WORKLOAD SPEEDUP WITH VALIDATOR AND MINER (WITHOUT MINING)

Workload (Average Speedup)		Total number of transactions executed per block				
		100	200	300	400	500
Serial Execution	Miner	1.00	1.00	1.00	1.00	1.00
	Validator	1.00	1.00	1.00	1.00	1.00
1 Slave	Miner	0.82	0.75	0.76	0.86	0.91
	Validator	0.78	0.71	0.69	0.82	0.89
2 Slaves	Miner	1.12	1.01	1.26	1.28	1.32
	Validator	1.03	1.05	1.22	1.24	1.45
3 Slaves	Miner	1.32	1.38	1.40	1.45	1.54
	Validator	1.05	1.34	1.56	1.65	1.72
4 Slaves	Miner	1.42	1.48	1.51	1.58	1.62
	Validator	1.21	1.28	1.31	1.52	1.72
5 Slaves	Miner	1.47	1.58	1.62	1.70	1.82
	Validator	1.28	1.29	1.36	1.79	1.89

The average speedup of miner and validator nodes for varying number of slave nodes as follows:

- 1 Slave: When using a single slave, both miners and validators experience a speedup of less than 1.00. This indicates that concurrent execution with only one slave is slower than sequential execution. The workload is not efficiently distributed among the participants.

- 2 Slaves: With two slaves, both miners and validators show a speedup greater than 1.00. This demonstrates that concurrent execution with two slaves is more efficient than sequential execution. The workload is better distributed, resulting in improved transaction throughput.
- 3 Slaves: As the number of slave increases to three, the speedup for both miners and validators further improves. This indicates that concurrent execution with three slaves is even more efficient in processing transactions compared to sequential execution.
- 4 Slaves: With four slaves, the speedup values continue to increase for both miners and validators, indicating a higher level of efficiency in parallel processing.
- 5 Slaves: The highest speedup values are observed when using five slaves for both miners and validators. This indicates that concurrent execution with five slaves offers the best transaction throughput improvement compared to sequential execution.

The results of the evaluation showcased the speedups achieved by miners and validators using the proposed scalable blockchain architecture. The architecture demonstrated significant improvements in transaction throughput compared to traditional serial execution. Additionally, network latency was reduced, indicating enhanced efficiency in block generation and transaction processing. To further emphasize the advantages and effectiveness of the proposed approach, a comparison with existing solutions was provided. The evaluation demonstrated that the Hybrid Shard Generation Algorithm and Static Analysis-based Data Partitioning technique outperformed conventional blockchain approaches, highlighting its potential to address scalability challenges effectively.

The results of the evaluation showcased the speedups achieved by miners and validators using the proposed scalable blockchain architecture. The architecture demonstrated significant improvements in transaction throughput compared to traditional serial execution. Additionally, network latency was reduced, indicating enhanced efficiency in block generation and transaction processing. To further emphasize the advantages and effectiveness of the proposed approach, a comparison with existing solutions was provided. The evaluation demonstrated that the Hybrid Shard Generation Algorithm and Static Analysis-Based Data Partitioning technique outperformed conventional blockchain approaches, highlighting its potential to address scalability challenges effectively. Overall, the results suggest that as the number of slaves increases in the system, the workload speedup also improves significantly. This demonstrates the benefits of parallel processing and how it enhances transaction throughput when validators and miners work concurrently in the blockchain network. The use of more slaves allows for better workload distribution and improved performance, contributing to the overall scalability and efficiency of the blockchain system.

VI. CONCLUSION

The proposed scalable blockchain architecture presents a significant step forward in addressing the critical challenge of scalability in blockchain networks. By leveraging the Hybrid Shard Generation Algorithm and Static Analysis-based Data Partitioning technique, the architecture effectively enhances transaction throughput and reduces network latency. The experimental evaluation using a blockchain simulation tool validates the effectiveness of these solutions, demonstrating notable improvements in performance. The results of the evaluation reveal that as the number of shards generated increases, transaction execution times by both miners and validators decrease, leading to improved transaction throughput. The speedup achieved by parallel mining over serial mining also increases with the rise in the number of shards, highlighting the advantages of parallel processing and efficient workload distribution. Moreover, the comparative study analysis emphasizes the significance of concurrent execution with multiple slaves, which substantially improves workload speedup for both miners and validators. This underscores the importance of parallel processing in achieving higher transaction throughput and overall network efficiency. The findings of the paper provide valuable insights into the potential of Hybrid Shard Generation and Static Analysis-Based Data Partitioning techniques in addressing scalability limitations in traditional blockchain platforms. These techniques offer practical solutions that make blockchain networks more efficient and adaptable for diverse applications in various industries. As research in this area continues to evolve, it is expected that these innovative techniques will play a pivotal role in shaping the future of blockchain networks, enabling them to handle increasing transaction volumes and meet the demands of an ever-evolving digital landscape.

REFERENCES

- [1] Dumitreloghin Ee-Chien Chang Hung Dang, Tien Tuan Anh Dinh. Towards scaling blockchain systems via sharding. In Proceedings of the 2019 International Conference on Management of Data. SIGMOD '19.
- [2] Deepal Tennakoon and Vincent Gramoli. Dynamic blockchain sharding. In 5th International Symposium on Foundations and Applications of Blockchain, 2022.
- [3] Marcos Antonio Vaz Salles Yizhong Liua, Jianwei Liua. Building blocks of sharding blockchain systems: Concepts, approaches, and open problems. In Sciencedirect, 2021.
- [4] Mark Nixon Song Ha Gang Wang, Zhijie Jerry Shi. Sok: Sharding on blockchain. In Proceedings of the 1st ACM Conference, Zurich, Switzerland, 2019. Advancement in financial technologies.
- [5] Mahnush Movahedi Mahdi Zamani and Mariana Raykova. Rapidchain: Scaling blockchain via full sharding. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, pages 931–948, 2018.
- [6] Praveen M Dhulavvagol, S G Totad, Performance Enhancement of Distributed System Using HDFS Federation and Sharding, Procedia Computer Science, Volume 218, 2023, Pages 2830-2841, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2023.01.254>.
- [7] G. Wood. Ethereum: a secure decentralised generalised transaction ledger. In Ethereum Project YELLOW paper, page 1–32, 2014.
- [8] S. Peri S. Rathor P. S. Anjana, S. Kumari and A. Somani. An efficient framework for optimistic concurrent execution of smart contracts. In 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pages 83–92, Feb, 2019.

- [9] O. Novo. Blockchain meets iot: an architecture for scalable access management in iot. In *IEEE Internet of Things Journal*, page 1184–1195, 2018.
- [10] S. Faust S. Dziembowski and K. Hostáková. General state channel networks. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, pages 949–966, Toronto, Canada, October, 2018.
- [11] Liuyang Ren and A. S. Ward. Transaction placement in sharded blockchains. Waterloo, Canada, 9 Jun, 2022. arXiv:2109.07670v3.
- [12] Jingjie Jiang Yuechen Tao, Bo Li. On sharding open blockchains with smart contracts. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, page 1357–1368, Dallas, TX, USA, 2020. IEEE.
- [13] M. Staples et al. X. Xu, I. Weber. A taxonomy of blockchain-based systems for architecture design. In *Proceedings of the 2017 IEEE International Conference on Software Architecture (ICSA)*, page 243–252, Gothenburg, Sweden, April 2017. IEEE.
- [14] Praveen M Dhulavvagol, Vijayakumar H Bhajantri, S G Totad, Blockchain Ethereum Clients Performance Analysis Considering E-Voting Application, *Procedia Computer Science*, Volume 167, 2020, Pages 2506-2515, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2020.03.303>.
- [15] Chuang Peng Runyu Chen, Lunwen Wang and Rangang Zhu. An effective sharding consensus algorithm for blockchain systems. In *Electronics*, Heifei, China, 2022.
- [16] Primicerio Lin, Jian-Hong and Kevin. Lightning network: a second path towards centralisation of the bitcoin economy. In *New Journal of Physics*, 2020.
- [17] K. Croman, C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, and E. G un. On scaling decentralized blockchains. In *Proc. 3rd Workshop on Bitcoin and Blockchain Research*, 2016.
- [18] J. Bonneau, A. Miller, J. Clark, A. Narayanan, J. A. Kroll, and E. W. Felten. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *2015 IEEE Symposium on Security and Privacy*, pages 104–121. IEEE, 2015.
- [19] PM Dhulavvagol, SG Totad, P Pratheek, Enhancing Transaction Scalability of Blockchain Network Using Sharding. In *Soft Computing for Security Applications: Proceedings of ICSCS 2023* 1449, 253
- [20] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *Proc. Secur. Privacy Work-shops (SPW)*, May 2015, pp. 180–184.
- [21] L. Luu, V. Narayanan, C. Zhang, K. Baweija, S. Gilbert, and P. Saxena. A secure sharding protocol for open blockchains. In *CCS*, 2016. [10] Koc et al., "Towards Secure E-Voting Using Ethereum Blockchain", 2018 IEEE 6th International Symposium on Digital Forensic and Security.
- [22] Suporn Pongnumkul, Chaiyaphum Siripanpornchana, Suttipong Thajchayapong, "Performance Analysis of Private Blockchain", 26th International Conference on Computer Communication and Networks (ICCCN), 2017.
- [23] M. M. Arer, P. M. Dhulavvagol and S. G. Totad, "Efficient Big Data Storage and Retrieval in Distributed Architecture using Blockchain and IPFS," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-6, doi: 10.1109/I2CT54291.2022.9824566.
- [24] Yue Hao, Yi Li, Xinghua Dong, Li Fang, Ping Chen, "Performance Analysis of Consensus Algorithm in Private Blockchain", *Intelligent Vehicles Symposium (IV) 2018 IEEE*, pp. 280-285, 2018.
- [25] T. T. A. Dinh, J. Wang, G. Chen, R. Liu, B. C. Ooi, K.-L. Tan, *Blockbench: A framework for analyzing private blockchains*.

Detection of Herd Pigs Based on Improved YOLOv5s Model

Jianquan LI¹, Xiao WU², Yuanlin NING³, Ying YANG^{4*}, Gang LIU⁵, Yang MI⁶

College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China^{1,2,3,4,5,6}

Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, Beijing 100083, China⁵

Key Laboratory of Agricultural Information Acquisition Technology, Ministry of Agriculture,
College of Information and Electrical Engineering, China Agricultural University, Beijing, China⁵

Abstract—Fast and accurate detection technology for individual pigs raised in herds is crucial for subsequent research on counting and disease surveillance. In this paper, we propose an improved lightweight object detection method based on YOLOv5s to improve the speed and accuracy of detection of herd-raised pigs in real-world and complex environments. Specifically, we first introduce a lightweight feature extraction module called C3S, then replace the original large object detection layer with a small object detection layer at the output (head) of YOLOv5s. Finally, we propose a dual adaptive weighted PAN structure to compensate for the information loss of feature map at the neck of YOLOv5s caused by down sampling. Experiments show that our method has an accuracy rate of 95.2%, a recall rate of 89.1%, a mean Average Precision (mAP) of 95.3%, a model parameter number of 3.64M, a detection speed of 154 frames per second, and a model layer count of 183 layers. Comparing with the original YOLOv5s model and the current state-of-the-art object detection models, our proposed method achieves the best results in terms of mAP and detection speed.

Keywords—Pig; deep learning; computer vision; object detection

I. INTRODUCTION

According to the data of the National Bureau of Statistics 0, the total meat production in China in 2022 was 92.27 million tons, of which the pork production was 55.41 million tons, accounting for about 60.05%. It can be seen that pig farming has become a pillar industry in China's livestock industry, and with the continuous expansion of pig farming, there is a growing demand for intelligent farming technologies to improve efficiency and productivity. Currently, the daily monitoring of pigs in herds is mainly performed by humans, which is subjective and time-consuming. Computer vision technologies such as object detection can realize automatic monitoring, which is more efficient and timely.

In recent years, many studies have investigated pig object detection. However, to obtain high detection accuracy in complex scenes such as piggery requires the deployment of large models, resulting in low detection speed and an inability to meet the real-time requirements. Furthermore, pigs tend to pile up and gather together, leading to severe occlusion and adhesion in images. In addition, pigs in images often appear as small objects occupying few pixels, which makes it more difficult to extract effective features and recognize. In

summary, the main challenges faced by the current pig detection based on computer vision are:

- Detection speed and accuracy cannot be balanced, that means higher detection accuracy need bigger model size with lower detection speed, and it is difficult to meet the demand of high accuracy and real-time detection simultaneously.
- The heavily occluded nature of pigs in real farming scenarios presents a challenge for increasing detection accuracy.
- Detecting small objects such as pigs that occupy few pixels in the image is difficult and often leads to missed detections.

To address the above problems, we propose an improved object detection algorithm based on YOLOv5s for group-farmed pig. The main contributions of this paper are summarized as follows:

- A lightweight module is introduced for achieving better performance and faster speed simultaneously. The C3 module in YOLOv5s with more branches and deeper model layers has a large number of parameters and computation, which affects the model detection speed. Therefore, a more lightweight C3S module is used instead of C3 module to improve the detection speed without reducing the accuracy.
- A small object detection branch is added to the detection layer. Since small objects account for a substantial part of the dataset, and the characteristics of small objects are difficult to detect. A small object detection branch is added instead of the large object detection layer to improve the detection capability of the model for small objects.
- A dual adaptive weighted PAN structure is proposed to enhance the feature extraction ability of the neck. In view of the complexity of the real farming environment, the dual adaptive weighted PAN structure can extract more feature for object detection and thus improve the detection accuracy of the model.

II. RELATED WORKS

In recent years, a series of achievements have been made in the field of individual pig detection based on deep learning.

The detection algorithms mainly consist of one-stage and two-stage, with the one-stage representative algorithms including the YOLO series [2], SSD [5], FOCs [6], etc., and the two-stage representative algorithms including RCNN [7], Fast-RCNN [8], Faster-RCNN [9], etc.

In terms of one-stage algorithms, Yan et al. [10] improved the detection accuracy without introducing additional computation by combining Tiny-YOLO with feature pyramid attention, achieving an accuracy of 85.85% on detecting pigs in group breeding. Shen et al. [11] used the YOLOv3 and FPN algorithm for detecting piglets, achieving a detection accuracy of 93.84%. Fang et al. [12] improved the CenterNet by using MobileNet as the feature extraction network to reduce the number of parameters and increase the computation speed. By introducing a feature pyramid structure to enhance the feature extraction ability, they achieved an mAP of 94.3%. Seo et al. [13] reduced the computational workload of 3×3 convolutions in YOLOv4 to achieve fast detection of individual pigs and improved accuracy through the generation of a three-channel composite image using simple image preprocessing techniques. Ahn et al. [14] combined the test results of two YOLOv4 models at the bounding-box level to increase the pig detection accuracy from 79.93% to 94.33%. These one-stage object detection algorithms have achieved satisfactory detection accuracy in scenarios with lower pig density and less occlusion and adhesion. However, in practical applications, they may not accurately reflect the desired performance. In real breeding conditions, there is still room for improvement in striking a balance between detection accuracy and speed.

In terms of second-stage algorithms, Riekert et al. [15] combined NAS (Neural Architecture Search) with the Faster-RCNN to detect the posture and position of pigs, achieving an average detection accuracy of 80.2%. Li et al. [16] used ResNet101 combined with the FPN algorithm as the backbone network and trained Mask R-CNN with transfer learning to detect pig crawling behavior, achieving a detection accuracy of 94.5%. However, their detection speed does not satisfy the real-time detection requirements, and their large model size makes them difficult to deploy on embedded devices.

III. DATASETS AND PROPOSED METHOD

A. Datasets

The data in this study were collected from Tianpengxingwang pig breeding farm in Shunyi, Beijing in November 2019. A Hikvision camera was fixed above the pigsty at an oblique angle to cover the entire pig pen. The video was recorded in MP4 format with a resolution of 1920×1080 and a frame rate of 30 frames per second. A frame was extracted from the collected video every two minutes or every 3600 frames, resulting in a total of 500 images.

To enrich the background and shooting angles of the dataset, a publicly available group-feeding pig dataset provided by iFlytek was added. This portion of the dataset contains a total of 920 images with a resolution of 1920 × 1080, a bit depth of 24 bits, and 3-channel RGB color images. The two parts of the dataset contain 1,420 images and 43,592 pigs in total. The shooting angle is from a top-down perspective, and the shooting time includes both day and night scenes. Table I

provides a statistical summary of pig density in the dataset used in this study. We augmented the images in the training set online, including HSV transformation, horizontal flipping, translation, proportional scaling, and Mosaic augmentation.

TABLE I. A STATISTICAL SUMMARY OF PIG DENSITIES

Number of individual pigs in a single image	Number of images
13-23	287
24-34	675
35-45	360
46-62	98

B. Improved YOLOv5s

In order to strike a balance between detection speed and accuracy, in this study, we consider adopting a one-stage approach as the base model. YOLOv5 [17] is a one-stage object detection model and has been improved on the basis of YOLOv4 [18], with the characteristics of small size, fast detection speed, and easy deployment. The improvement points that greatly improve its speed and accuracy mainly include the following four aspects:

- Input: Mosaic data augmentation, adaptive anchor box calculation and adaptive image scaling.
- Backbone: CSP structures, Focus structure.
- Neck: The Path Aggregation Network (PAN) [19] and Feature Pyramid Network (FPN) [20] structures are added between the backbone and head layers as a neck network.
- Head: Loss function CIOU-Loss [21] during train, IOU [22] during prediction and Non-maximum suppression (NMS) for prediction box screening.

There are four versions of YOLOv5, which are YOLOv5s, YOLOv5m, YOLOv5l and YOLOv5x, with the network depth and width increasing progressively. While the larger versions of YOLOv5 have higher detection accuracy, they also have a larger number of parameters and computation, making them less suitable for real-time detection scenarios. As a result, we have chosen to use YOLOv5s as the base network, as it has the smallest number of parameters among the four versions and can provide a reasonable trade-off between detection accuracy and computational efficiency. However, it still cannot fully meet the demands of detection of herd pigs, which require faster and more accurate object detection algorithms.

To address the challenges posed by severe occlusion and aggregation of pigs in group-raised pig farming scenarios and the requirement for real-time detection, we propose an improved version that significantly enhances its performance. Fig. 1 shows the structure details of improved YOLOv5s. By replacing C3 with C3S module, the number of parameters and calculation amount are greatly reduced, and the detection speed is accelerated while the accuracy remains unchanged. We also add a small object layer to improve the ability of the network to detect small objects. Finally, a dual adaptive weighted PAN structure is proposed to enhance the feature extraction ability and further improve its detection accuracy.

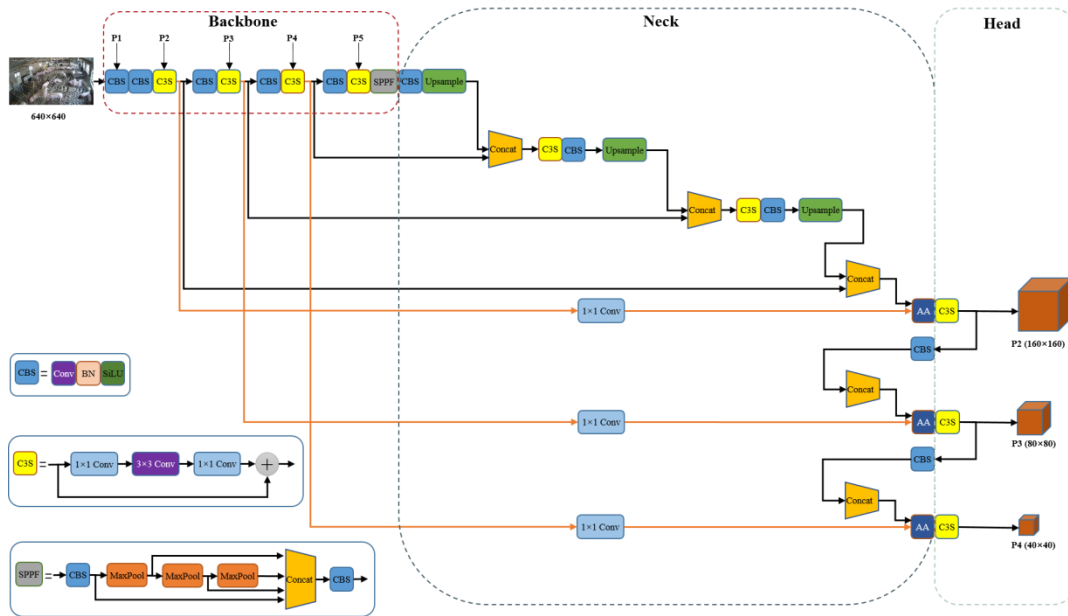


Fig. 1. The structure of improved YOLOv5s.

C. C3S Module

C3 module consists of convolutional layers (Conv), batch normalization layers (BN), the SiLU activation function, addition operations (add), and feature map concatenation along the channel dimension (Concat). The depth factor controls the number of BottleNeck modules in the structure and can be adjusted to control the model's depth. While the C3 module improves detection performance, it can also result in a deep model that reduces inference speed and increases computational cost and parameter count.

To improve the efficiency of the model, we introduced a lightweight convolutional module called C3S, which replaced the original C3 module. In Fig. 1, the C3S module is composed of 1x1 convolutions, 3x3 convolution, and residual structure to enhance the model's expressive power and feature extraction ability. Specifically, the 1x1 convolution can not only reduce the channel dimension but also promote inter-channel information exchange. Therefore, we first use 1x1 convolutions to reduce the input feature map's channel number by half, aiming to decrease the parameter count and computation cost. Next, we incorporate 3x3 convolutions to strengthen the feature extraction ability, increasing the feature map's channel number to twice the current number and subsequently using another 1x1 convolution to perform cross-channel information integration. Finally, we introduce residual structure to prevent gradient vanishing or explosion during model training. To further improve efficiency, we also decrease the channel number of the C3S module to 3/4 of the original by controlling the width factor.

D. Small Object Detection Layer

The input image resolution is 640x640, and then the downsampling operation is used by convolution with a stride of two, resulting in output feature maps with half the width and height of the input feature map. P1, P2, P3, P4, and P5 denote feature maps obtained via convolutional layers with

downsampling steps of 1, 2, 3, 4, and 5, respectively, resulting in resolutions of 320x320, 160x160, 80x80, 40x40, and 20x20. As shown in Fig. 2, the head of YOLOv5s model consists of three detection layers, which take input feature maps of different resolutions (P3, P4, and P5). The P3 feature map, with the lowest resolution, is used to detect small objects, while the P4 and P5 feature maps are used to detect medium and large objects, respectively. In our dataset, small objects account for a large proportion of the total. Therefore, we replace the original large object detection layer in the head with a smaller one (as indicated by the red solid line) to enhance the model's ability to detect small objects in the images. The removed module is indicated by the light green dashed line.

Directly upsampling the feature map results in four times computational cost increase, which can negatively impact the inference speed. To address this issue, we first reduce the dimension of the input feature map before upsampling, which can partially alleviate the increase in parameters and computational cost. This approach improves the model's inference speed compared to direct upsampling.

$$\begin{aligned}
 Calculated = & B \times O \times [C \times (\frac{H-K+P_h}{S} + 1) \times (\frac{W-K+P_w}{S} + 1) \times (2 \times K \times K - 1) \\
 & + (C-1) \times (\frac{H-K+P_h}{S} + 1) \times (\frac{W-K+P_w}{S} + 1)]
 \end{aligned} \tag{1}$$

Eq. (1) represents the calculation formula for the convolutional computation. In this equation, *Calculated* represents the convolutional computation, *B* represents the batch size used during training, *O* represents the number of input feature map channels, *C* represents the number of output feature map channels, *H* and *W* represent the height and width of the input feature map, *P_w* and *P_h* represent the number of pixels padded in the height and width directions, respectively, *S* represents the stride of the convolutional kernel and *K* represents the size of the convolutional kernel (*H* and *W* are much larger than *K*, *S*, *P_w* and *P_h*).

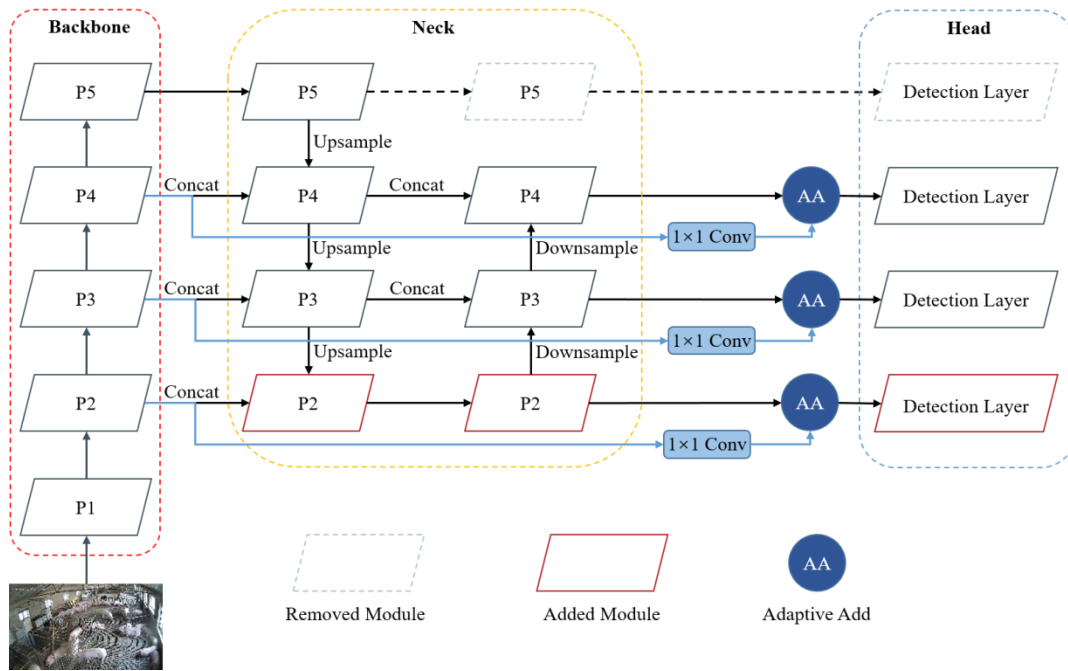


Fig. 2. Sketch of the model structure after adding the small object detection layer and dual adaptive weighted PAN structure.

E. Dual Adaptive Weighted PAN Structure

A dual adaptive weighted PAN structure was proposed to enhance the feature extraction ability of the neck. The blue part in Fig. 2 represents the dual adaptive weighted PAN structure, which consists of 1×1 convolutions and adaptive addition (AA) operations. In the YOLOv5s model, the features extracted by the backbone are fed into the neck for feature fusion. The neck performs multiple downsampling and upsampling operations to generate feature maps of different sizes for detecting objects of different sizes. However, some feature information is inevitably lost during the downsampling process. To address this issue, we reuse the original features extracted by the backbone and adjust their channel numbers using 1×1 convolutions to match the channel numbers of the large, medium, and small object feature maps output by the neck. We then perform adaptive addition between the adjusted feature maps and the object feature maps. Since the importance of backbone features and neck features may not be the same, direct addition may assume equal importance. Therefore, we define this addition operation as adaptive weighted addition, where a learnable weight is used to adjust the importance of the two types of features. We train the model using backpropagation and gradient descent to update the weight until convergence is reached.

$$X_{out} = w \times X_{input1} + (1-w) \times X_{input2} \quad (2)$$

Eq. (2) represents the adaptive add operation, where X_{out} represents the output feature map, w represents the weight, X_{input1} represents input feature map A, and X_{input2} represents input feature map B.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Parameters Setting

The experiments in this study were conducted in Linux Ubuntu 18.0.4 environment (CPU: Inter Core i9 10900K, GPU: Nvidia GeForce RTX3060 \times 2, RAM: 64G), and the deep learning framework was Pytorch 1.9.0. See Table II for other parameter settings. The experiments in this paper were conducted using the same experimental configuration.

TABLE II. PARAMETERS SETTING IN THIS PAPER

Configuration	Value
Optimizer	SGD
Learning Rate	0.01
Momentum	0.937
Weight Decay	0.0005
Batch Size	8
Training Epochs	200

B. Evaluation Metrics

The evaluation metrics used in this paper are precision, recall, mAP, model depth, parameter count, computational complexity, F1 score, and Frames Per Second (FPS). Precision represents the proportion of true positive predictions among all positive predictions made by the model, while recall represents the proportion of true positive predictions among all actual positive instances. The mAP is related to both precision and recall, with a higher mAP indicating a higher average detection accuracy of the model. The model's parameter count, computational complexity, and depth affect the inference speed of the model, while FPS measures the number of images the model can process per second, indicating the computational speed of the model. The F1 score is the harmonic mean of

precision and recall and is used to measure the overall performance of the model. The equations for calculating these evaluation metrics are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$AP = \int_0^1 P dR \quad (5)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (6)$$

$$F1 = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (7)$$

In the above equations, N represents the total number of classes and TP , FP and FN represent the number of true positive predictions, false positive predictions and false negative predictions, respectively. P and R are abbreviations for precision and recall, respectively.

C. Impact of Different Detection Layers on Model Performance

To explore the impact of varying numbers of detection layers and input feature map resolutions on model performance, we performed three sets of comparative experiments, and the results are presented in Table III. Experiment 1 used YOLOv5's original detection layers ([P3, P4, P5]). Experiment 2 added a small object detection layer with input resolution P2 ([P2, P3, P4, P5]). Compared to Experiment 1, although the precision decreased by 0.4%, the recall and mAP increased by 1.7% and 0.7%, respectively, in Experiment 2. Therefore, the overall performance of the model with the added small object detection layer was superior to that of the original YOLOv5 model. Experiment 3 removed the large object detection layer with input resolution P5 ([P2, P3, P4]). Compared to Experiment 2, the precision, recall, and mAP of the model further improved in Experiment 3, with precision increasing by 0.4%, recall increasing by 0.7% and mAP increasing by 0.5%.

In summary, the performance of the model was improved by adding a small object detection layer, as the feature map with input resolution P2 contained more information about small objects, making it easier to detect them. Furthermore, removing the large object detection layer with input resolution P5 led to further improvements in the model's performance. We believe this is because the number of large objects in our dataset was relatively small, resulting in fewer positive samples allocated to the large object detection layer during training. As a result, the parameters of the large object detection layer were difficult to optimize, making it challenging to accurately predict the presence of large objects, which ultimately affected the overall detection accuracy.

TABLE III. IMPACT OF DIFFERENT DETECTION LAYERS ON MODEL PERFORMANCE

Experiment Number	Input of detection layer Precision(%)	Precision (%)	Recall (%)	mAP (%)
1	[P3, P4, P5] ^a	94.8	85.3	93.3
2	[P2, P3, P4, P5]	94.4	87	94
3	[P2, P3, P4]	94.8	87.7	94.5

^a The input of the detection layer [P3, P4, P5] indicates that there are 3 detection layers, and the input resolutions are 8, 16, and 32 times downsampled from the input image, respectively.

D. Comparison with Different Object Detection Models

To verify the superiority of the improved YOLOv5s model proposed in this paper for individual pig detection in group feeding, we compared its performance with the other five common object detection models, such as Faster-RCNN [9], CenterNet [23], YOLOv3 [4], YOLOv4 [18], and YOLOX [24]. The experiments were conducted using the same experimental configuration. Fig. 3 shows the training accuracy curves for the six models, with the horizontal axis representing Epoch and the vertical axis representing mAP values. It can be seen that our method achieved the highest accuracy during the training phase.

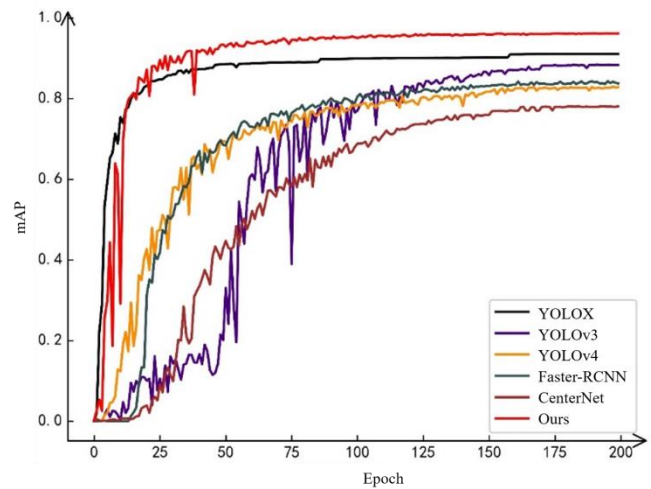


Fig. 3. Training accuracy curves of different models.

Table IV illustrates the detailed quantitative analysis of the model performance on the test dataset under same experimental configuration. It can be found that our proposed method outperforms other methods in terms of accuracy, speed, and computational efficiency. The mAP of our method achieved the highest value of 95.3%, while the FPS reached 154 frames/second. The computational complexity and number of parameters were also the lowest, with only 16.7 GFLOPs and 3.64M parameters, respectively.

To provide a more intuitive comparison of the performance of different models for individual pig detection in group feeding, Fig. 4 compares the detection results of the improved YOLOv5s model with those of other models. The green arrows in the figure indicate the missed pigs detected by each model. Compared to our model, the other models all showed missed detections, and the missed pigs in these models were all in

heavily occluded, with two of them showing only a small part of their bodies, which is a small object detection problem. These results demonstrate that the improved YOLOv5s model performs better in detecting occluded and sticky pigs, as well as small objects, effectively improving the accuracy of individual pig detection in group feeding.

E. Ablation Study

To explore the effectiveness of the proposed improvements in this paper, we conduct extensive ablation study of C3S module, dual adaptive weighted PAN structure and small object detection layer. Table V shows that the C3S module significantly reduces the number of parameters, computation

cost, and model layers while maintaining almost the same detection accuracy, demonstrating its effectiveness in improving the detection speed. Additionally, the proposed dual adaptive weighted PAN structure improves the precision, recall, and mAP of the model while keeping the computation cost, parameter count, and detection speed almost unchanged. By improving the detection layer, the model's recall and mAP are effectively increased with minimal reduction of detection speed. The improved YOLOv5s model shows a 3.8% increase in recall and a 2% increase in mAP, while reducing the parameter count by 48%, increasing FPS by 12.4%, and reducing model depth by 22%.

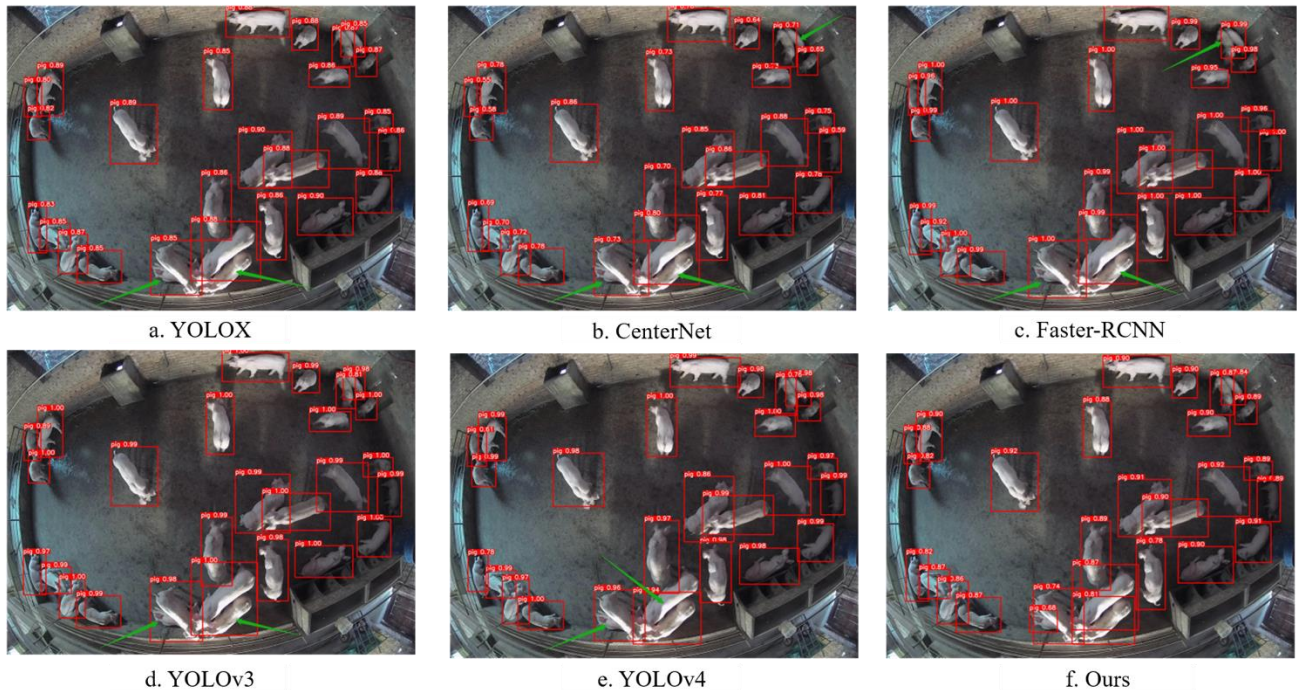


Fig. 4. Comparison of the detection effect of different models.

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT OBJECT DETECTION MODELS

Model	Backbone	mAP (%)	Params (MB)	FLOPs(G)	FPS	Model size (MB)
Faster-RCNN	VGG16	86.28	136.7	369.9	20	521
YOLOv3	DarkNet53	92.19	61.5	65.52	47	235
YOLOv4	CSPDarkNet53	86.55	63.9	60	46	244
YOLOX	CSPDarkNet53	93.06	8.9	26.6	61	34
CenterNet	ResNet50	86.83	32.7	70	53	125
Ours	CSPDarkNet53	95.3	3.64	16.7	154	8

TABLE V. RESULTS OF ABLATION EXPERIMENT

Model	Precision (%)	Recall (%)	mAP (%)	Parameters (MB)	FLOPs (G)	FPS	Layers
YOLOv5s	94.8	85.3	93.3	7	15.8	137	235
YOLOv5s+a	94.7	84.8	93	4.4	9.5	169	150
YOLOv5s+a+b	95.3	87.5	94.1	4.57	9.9	164	168
YOLOv5s+a+b+c ^b	95.2	89.1	95.3	3.64	16.7	154	183

^b: Note: a is the C3S module, b is the dual adaptive weighted PAN structure, and c is the small object detection layer.

F. Compared to Existing One-stage Pig Detection Methods

Table VI shows a comparison of the detection results achieved by our proposed improved YOLOv5s model on the test set, as well as the reported results of existing one-stage herd pig detection methods. Our dataset has a higher average number of pigs per image compared to the data reported in [10], [12] and [14]. Moreover, the pigs in our dataset exhibit higher levels of occlusion and adhesion, which increases the difficulty of object detection. Compared to the methods proposed in [10], [12] and [14], our proposed method achieved improvements of 9.45%, 1% and 0.97%, respectively, in terms of mAP.

TABLE VI. COMPARISON OF IMPROVED YOLOV5S AND EXISTING ONE-STAGE PIG DETECTION METHODS

Method	Average number of pigs per image	Model	mAP (%)
[10]	6	FPA-Tiny-YOLO	85.85
[12]	13	MF-CenterNet	94.30
[14]	9	YOLOv4	94.33
ours	31	Improved YOLOv5s	95.30

V. CONCLUSION

In this paper, an improved lightweight object detection method is proposed based on YOLOv5s, which has achieved higher accuracy and faster detection speed in high pig density scenarios with severe occlusion and adhesion. The lightweight C3S module proposed in this paper reconstructs the backbone and neck, resulting in a significant reduction in model parameters, computational complexity, and model depth. These modifications greatly enhance the detection speed to meet the requirements of real-time detection. The proposed dual adaptive PAN structure enhances the feature fusion capability of the neck, leading to improved detection accuracy. Furthermore, replacing the original large object detection layer with a small object detection layer in the detection stage significantly increases the recall rate and average detection precision.

Compared to existing methods for herd pig detection, our method simultaneously satisfies the requirements of high accuracy and real-time detection, making it deployable in practical group-raised farming scenarios and providing significant technical support for disease monitoring and pig counting. For future work, we expect to further enhance the feature extraction capabilities of the backbone network and streamline the model to construct a faster and more accurate object detection model for group-raised pig monitoring.

ACKNOWLEDGMENT

This work was supported by National Key R&D Program of China (Grant No. 2021YFD1300502).

REFERENCES

[1] National Bureau of Statistics of the People's Republic of China. China Statistical Yearbook [J]. Peking: China Statistics Press, 2022.
[2] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[3] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
[4] Redmon J, Farhadi A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv:1804.02767, 2018.
[5] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
[6] Tian Z, Shen C, Chen H, et al. Fcos: Fully convolutional one-stage object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9627-9636.
[7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
[8] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[9] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. Advances in neural information processing systems, 2015, 28.
[10] Hongwen Y, Zhenyu L, Qingliang C. Multi-target detection based on feature pyramid attention and deep convolution network for pigs [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(11): 193-202.
[11] SHEN M, TAI M, CEDRIC O. Real-time detection method of newborn piglets based on deep convolution neural network [J/OL][J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8): 270-279.
[12] Fang J, Hu Y, Dai B, et al. Detection of group-housed pigs based on improved CenterNet model [J]. Trans. Chin. Soc. Agric. Eng, 2021, 37: 136-144.
[13] Seo J, Ahn H, Kim D, et al. EmbeddedPigDet—Fast and accurate pig detection for embedded board implementations[J]. Applied Sciences, 2020, 10(8): 2878.
[14] Ahn H, Son S, Kim H, et al. EnsemblePigDet: Ensemble deep learning for accurate pig detection[J]. Applied Sciences, 2021, 11(12): 5577.
[15] Riekert M, Klein A, Adrion F, et al. Automatically detecting pig position and posture by 2D camera imaging and deep learning[J]. Computers and Electronics in Agriculture, 2020, 174: 105391.
[16] Li D, Zhang K, Li X, et al. Mounting behavior recognition for pigs based on Mask R-CNN[J]. Trans. Chin. Soc. Agric. Mach, 2019, 50: 261-266.
[17] Jocher G, Chaurasia A, Stoken A, et al. ultralytics/yolov5: v6. 1-tensorrt, tensorflow edge tpu and opencv export and inference[J]. Zenodo, 2022.
[18] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[19] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
[20] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
[21] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
[22] Yu J, Jiang Y, Wang Z, et al. Unitbox: An advanced object detection network[C]//Proceedings of the 24th ACM international conference on Multimedia. 2016: 516-520.
[23] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.
[24] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

The Impact of Cyber Security on Preventing and Mitigating Electronic Crimes in the Jordanian Banking Sector

Tamer Bani Amer¹, Mohammad Ibrahim Ahmed Al-Omar²

Assistant Professor, Department of Computer Science, Jadara University, Irbid, Jordan¹

Assistant Professor, Department of Computer Networks and Cyber Security, Jadara University, Irbid, Jordan²

Abstract—As technology advances and cyber threats continue to evolve, cyber security professionals play a critical role in developing and implementing robust security measures, staying ahead of potential risks, and mitigating the impact of cyber incidents. Many studies have examined the impact of cyber security on banks, without focusing on electronic crimes. Despite its importance, to the best of our knowledge, there are no studies on the impact of cyber security on mitigating electronic crimes in the banking sector. Therefore, the purpose of this study is to ascertain how cyber security affects electronic crimes in the Jordanian banking industry. The study sample consisted of 270 senior Jordanian managers and employees who understand the importance of cyber security in the banking sector in 14 Jordanian commercial banks, listed on the Amman stock exchange. The study used SPSS to evaluate how banks can enhance network security infrastructure to prevent unauthorized access and data breaches and also to find out the role of cybersecurity in granting competitive advantage to banks. A relative importance index (RII) was conducted to rank the importance of variables' statements and test the hypotheses. The results found the most important method through which banks can effectively mitigate the risk of electronic crimes and ensure the security of customers' financial data is that banks utilize robust encryption technologies to ensure the protection of customer financial data while it is being transmitted and when it is stored (RII=0.740). About 81.5 % of the sample agree, also, banks that have a strong cyber security system provide a secure platform for digital financial services which increases the competitive advantage as they were ranked first for their relative importance at both the category level and overall ranking with (RII=0.754). The study recommended that the banking industry, must consistently educate its customers on information security techniques and how to avoid hacking into their accounts, and develop an alert system that can raise awareness for both banks and bank customers if there is any possible entry or access to the customer's account or organization confidential information.

Keywords—Cyber security; electronic crime; Jordanian banks; banking sector

I. INTRODUCTION

As technology advances and cyber threats continue to evolve, cyber security professionals play a critical role in developing and implementing robust security measures, staying ahead of potential risks, and mitigating the impact of cyber incidents. Moreover, cyber security refers to the practice of protecting computer systems, networks, and data from

unauthorized access, use, disclosure, disruption, modification, or destruction [1]. With the increasing reliance on technology and the interconnectedness of digital systems, cyber threats have become a significant concern for individuals, organizations, and governments worldwide [2]. The field of cyber security encompasses various measures, technologies, and practices designed to defend against cyber threats and ensure the confidentiality, integrity, and availability of information [3]. It involves protecting not only computers and servers, but also other devices connected to networks, such as smartphones, tablets, and the Internet.

Electronic crimes refer to illegal activities that are carried out using electronic devices, networks, or the internet [4]. With the rapid advancement of technology and the increasing reliance on digital systems, electronic crimes have become a significant concern for individuals, businesses, and governments worldwide [5]. In addition, to combat electronic crimes, governments, law enforcement agencies, and cyber security professionals work together to develop stringent laws, improve cyber security measures, raise public awareness, and promote digital literacy [6]. It is crucial for individuals and organizations to take necessary precautions, such as using strong passwords, keeping software up to date, and practicing safe online behavior, to protect themselves from falling victim to electronic crimes.

According to Arcuri et al. [7], the median cost of electronic crimes has climbed by approximately 200 percent in the last five years. Electronic crime expenditures quadrupled between 2015 and 2019, and it appears that they will double again between 2019 and 2024 [8]. Nonetheless, a considerable fraction of electronic crimes go unnoticed, such as industrial espionage getting access to confidential information. According to Seete [9], cyber risk is a huge potential threat to public and private institutions because of its effects on organizational information systems, reputation, loss of stakeholders' confidence, and financial losses. According to Bonfanti [10], the daily operations of virtually every person and organization are impacted by cyber security. Therefore, we must try to protect ourselves, our customers, and the supply chain against the loss of personal or sensitive information as technology enhances the flexibility, agility, and global reach of day-to-day operations. Must also be on the lookout for intellectual property theft, brand or reputation damage, and of course, monetary and economic losses.

Accordingly, the current study questions can be represented as follows questions:

- How can banks effectively mitigate the risk of electronic crimes and ensure the security of customers' financial data?
- How banks can enhance network security infrastructure to prevent unauthorized access and data breaches?
- How strong cyber security can give banks a competitive advantage?
- How banks can effectively educate and train employees and customers to be aware of and prevent cyber threats?

Many studies have examined the impact of cyber security on banks, without focusing on electronic crimes. However, the results are mixed [11, 12, 13] with cyber security having both a positive and a negative impact. Despite its importance, to the best of our knowledge, there are no studies on the impact of cyber security on electronic crimes in the banking sector. Therefore, the purpose of this study is to ascertain how cyber security affects electronic crimes in the Jordanian banking industry. In the unique context of the Jordanian banking industry, the study offers actual proof of the link between cyber security measures and the prevention and mitigation of electronic crimes. This empirical finding adds to the body of knowledge already available in digital crime prevention and cyber security. The study may aid in the creation and improvement of theoretical frameworks that clarify the complex link between cyber security measures and the decline in electronic crimes. This could improve the theoretical underpinnings of studies in cyber security. By concentrating on the Jordanian banking sector, the study provides context-specific insights into a particular business. For researchers, policymakers, and practitioners looking to grasp the complexities of cyber security within financial institutions, these sector-specific results can be helpful. The research will be separated into sections. We present a literature review and hypothesis development in Section II. We describe the data and methodology in Section III. We review the findings in Section IV and offer our discussions in Section 5 and conclusions in Section 6.

II. LITERATURE REVIEW AND HYPOTHESES DEVELOPMENT

A large number of studies [14, 15, 16, 17, 18, 19, 20, 21] deal with information security breaches, but there is still a limited amount of literature related to the banking sector. Electronic crimes have an unknown economic impact. A breach in information security has a negative economic impact, including decreased sales revenues, more expenses, a loss in future profits and dividends, a deterioration in reputation, and a decrease in market value [22]. Market value symbolizes investor confidence in a bank, and evaluating it is one way to calculate the damage of electronic crimes. Furthermore, Arcuri et al. [7] claim that investor behavior is influenced by what they have seen in the past, i.e., investors make decisions based on the impact of security breaches on a bank's market value in the past.

In the same context, information security plays a crucial role in combating electronic crimes [23]. Electronic crimes, also known as electronic crimes, encompass a wide range of illegal activities that are carried out using digital technologies and the internet. According to Wang et al. [12], these crimes include hacking, identity theft, phishing, malware attacks, data breaches, online fraud, and more. According to Mughal [24], Information security measures such as strong passwords, encryption, access controls, and multifactor authentication help prevent unauthorized access to sensitive data and systems. In addition, information security ensures the integrity and confidentiality of data [25]. Encryption techniques, secure data transmission protocols, and secure storage mechanisms help safeguard data from unauthorized modification, interception, or disclosure. Maintaining data integrity and confidentiality is vital in protecting sensitive information from cybercriminals who exploit it for financial gain or other malicious purposes [26].

According to Burton-Howard [27], network security plays a crucial role in mitigating electronic crimes in firms. Furthermore, effective network security measures help protect an organization's digital assets, including sensitive data, intellectual property, and financial information, from unauthorized access, data breaches, and cyberattacks [28]. Network security measures, such as firewalls, intrusion detection systems (IDS), and access controls, help prevent unauthorized individuals from gaining access to a firm network and systems [29]. Network security plays a vital role in safeguarding sensitive data from unauthorized disclosure or manipulation [30]. According to Prasad et al. [31], network security tools and technologies, such as intrusion prevention systems (IPS), security information and event management (SIEM) systems, and advanced threat detection solutions, enable firms to detect and respond to electronic crimes more effectively. According to Rafea et al. [32], network security plays a critical role in mitigating electronic crimes in the banking sector. As technology has advanced, so have the methods and sophistication of cybercriminals. By implementing robust access controls and encryption protocols, banks significantly reduce the risk of cybercriminals gaining unauthorized access to sensitive financial information [6]. According to Zainal et al. [33], banks store vast amounts of personal and financial information about customers, including account details, social security numbers, and transaction records.

In the same context, operational security known as OPSEC, plays a crucial role in mitigating electronic crimes [34]. According to Bandari [35], operational security measures, such as strong access controls, authentication protocols, and encryption, help prevent unauthorized individuals from gaining access to critical sensitive customer data. Banks hold vast amounts of personal and financial information about customers, making them attractive targets for cybercriminals [36]. According to Ritchot [37], effective operational security measures safeguard this data, ensuring its confidentiality, integrity, and availability. By implementing robust data encryption, secure storage practices, and regular data backups, banks mitigate the risk of data breaches, identity theft, and financial fraud. Operational security encompasses

advanced monitoring and detection systems that identify suspicious activities or anomalies within banking networks [38]. By deploying intrusion detection and prevention systems, security information and event management (SIEM) tools, and real-time monitoring solutions, banks promptly detect electronic crimes, such as malware infections, phishing attempts, or network intrusions [39]. Rapid detection allows for swift incident response, minimizing the potential damage caused by cybercriminals. According to Rivaldo et al. [40], maintaining a strong reputation and customer trust is vital for banks. Operational security measures, including robust cybersecurity frameworks and transparent communication about security practices, contribute to building customer confidence [41].

End-user education plays a crucial role in mitigating electronic crimes in the banking sector [42]. According to Catota et al. [43], end-user education, such as bank-customers and employees, about the risks, best practices, and preventive measures, the overall security posture of the banking sector be significantly improved. According to Alkhalil et al. [44], phishing attacks are a common method used by cybercriminals to trick users into revealing sensitive information like passwords, credit card numbers, or social security numbers. Through education, end-users learn to identify phishing attempts, recognize suspicious emails or websites, and avoid falling victim to such scams. This knowledge helps protect personal and financial information from being compromised. End-user education about safe online practices, such as keeping software and devices updated, avoiding suspicious downloads or attachments, and using secure networks, helps prevent malware infections and unauthorized access to sensitive information [45]. In the same context, end-user education and report suspicious activities promptly. Prompt reporting of potential electronic crimes to the bank security teams helps prevent further damage and enables the bank to take appropriate measures to investigate and mitigate the threat [46]. Reporting incidents also aids in the identification of emerging trends and the development of proactive security measures. Based on the aforementioned, the study develops the following hypotheses:

H1. Cybersecurity plays a crucial role in preventing and mitigating electronic crimes.

H2. Improving network security infrastructure contributes to preventing unauthorized access and data breaches.

H3. Strong cyber security practices give banks a competitive advantage.

H4. Training and educating employees and customers contribute effectively to preventing electronic crimes.

III. RESEARCH METHODOLOGY

A. Research Population and Sampling

The commercial banks listed on the Amman Stock Exchange will be the subject of the research. As the research sample, all 14 Jordan Commercial banks listed on the ASE were chosen. 270 personnel and managers who worked in various areas of the commercial banks listed on the ASE were the research participants. Due to the challenges in precisely

identifying the research population, convenience sampling, a non-probability sample technique, was used. To ascertain the impact of cyber security on electronic crimes in the Jordanian banking sector, the research looks at a questionnaire.

B. Research Design

Based on feedback from bank managers, literature research, and pre-survey analysis, the questionnaire will be created. To make it simpler for the participants to understand, it will be divided into three sections. Demographic information about the participants, such as age, gender, years of experience, position, and educational background, was requested in the first section of the survey. The second half concentrated on cyber security in commercial banks. The dependent variable (Electronic Crimes) was the subject of the third section measurement. A five-point Likert scale, with a rating of 1 for least important and 5 for most important, was used in the questionnaire. The selected sample consisted of senior Jordanian managers and employees who understood the importance of cyber security in the banking sector.

C. Measurement of Research Variables

The current research included several variables that required careful measurement to test the hypotheses and produce useful results. Cybersecurity was used as an independent variable. The electronic crimes in Jordanian banks served as the dependent variable. A structured online questionnaire and key performance indicators will be used to measure the variables. Accordingly, the data was analyzed, and the mean and standard deviation were determined.

To measure the validity of the study, the questionnaire was presented to a group of experts and specialists in the field of cybercrime and cybersecurity to review the data and receive their comments to ensure the suitability of the items of the questionnaire. The questionnaire was written in both Arabic and English to ensure high participation in this work and to obtain diverse perceptions from the sample. The results were analyzed using the SPSS statistical package for social sciences, descriptive analysis, which mainly used the factors of frequency, percentage of the study, and relative importance. After confirming the validity of the study tool and obtaining permission from the sample. A total of 300 questionnaires were distributed and 270 correct answers were received with a response rate of 90%. Responses with missing data or that did not meet serious data were excluded from the analysis.

The study examined the reliability of significant variables widely used in social studies. The main objective of this test is to verify the reliability of the items measuring variables to measure the target factors, also called internal consistency [47]. Cronbach Alpha is the most commonly used measure to perform a reliability analysis of the validity of measurement items [48], the reliability coefficient is rated between 0 to 1, although different assumptions have discussed this issue and suggest different acceptable values, the higher the value of the coefficient, the higher the degree of reliability of the measurements, the reliability analysis for this study was 0.885 which is high.

The feedback from the respondents has been analyzed and the Relative Importance Index technique was used for ranking.

The five-point Likert scale ranging from 1 (very low important) to 5 (very high important) was transformed into Relative Importance Index (RII) for each variable. The RII value has a range from 0 to 1 (0 not inclusive) and has been categorized into five levels of importance as shown in Table I.

TABLE I. RELATIVE IMPORTANCE INDEX VALUE

RII value	Importance level
From 0.8 to 1	High (H)
From 0.6 to 0.8	High-Medium (H-M)
From 0.4 to 0.6	Medium (M)
From 0.2 to 0.4	Medium-Low (M-L)
From 0 to 0.2	Low (L)

IV. RESULTS

A. Demographics Characteristics of Participant

The study includes 270 senior Jordanian managers and employees who understand the importance of cyber security in the banking sector, for gender distribution, the percentage of males 186 (68.9%) was higher than the percentage of females 84 (31.1%), and most of them were aged (31-40) while 94 (34.8%) of them more than 40 years old and 64 (23.7%) of them between (25 -30) About years of experience, the experience of the majority of the sample ranged from (1-5) years, at a rate of 37.4% and 78 (28.9%) from (6-15) years,61(22.6%) from (16-25) and 30 (11.1%) of them from 26-35 years. The majority of the sample was 57.8% of employees and 42.2% of managers. Table II summaries the demographic characteristics of the participants.

TABLE II. RESPONDENTS' DEMOGRAPHICS

Demographic	Frequency	Percent
Gender		
Male	186	68.9
Female	84	31.1
Age		
25-30	64	23.7
31-40	112	41.5
More than 40	94	34.8
Years of experience		
1-5	101	37.4
6-15	78	28.9
16-25	61	22.6
26-35	30	11.1
Nature of work		
Manager	114	42.2
Employee	156	57.8

The results about the normality distribution of the data showed a normally distributed dataset with a range of ± 1.00 to ± 2.00 of the normality distribution measure of skewness and kurtosis respectively. Related to the first question of the study, which states “How can banks effectively mitigate the risk of electronic crimes and ensure the security of customers’ financial data?” To answer this question, we analyze the results of cyber security in commercial banks as the

independent variable and ensuring network security infrastructure and mitigating risks of electronic crimes in banks, SPSS was used to calculate the mean of distribution and the standard deviation of each statement. Table III summarizes the results.

TABLE III. ARITHMETIC MEANS AND STANDARD DEVIATIONS ARE THE ESTIMATES OF THE STUDY SAMPLE ON ENSURING NETWORK SECURITY INFRASTRUCTURE AND MITIGATING RISKS OF ELECTRONIC CRIMES IN BANKS

No.	Statement	Mean	SD
1	Banks regularly provide comprehensive training programs to their employees on electronic crimes and protocols for cybersecurity.	3.04	1.113
2	Banks educate their customers on the risks of electronic crimes, cybersecurity, and the best practices for safeguarding their financial information.	3.23	0.964
3	Banks employ multi-factor authentication methods, such as hardware tokens and SMS-based verification, for customer logins.	3.23	0.925
4	Banks embrace biometric authentication techniques like fingerprint and facial recognition to verify the identity of customers.	3.32	1.136
5	Banks utilize robust encryption technologies to ensure the protection of customer financial data while it is being transmitted and stored.	3.70	0.806
6	Banks enforce strong password policies for both employees and customers.	3.30	0.647
7	Banks conduct regular security audits and vulnerability assessments of their IT systems.	3.49	0.865
8	Banks employ real-time fraud detection and prevention systems.	3.43	1.028
9	Banks have well-defined incident response plans and robust systems in place to recover from electronic crimes.	3.43	1.195
10	Banks implement systems that continuously monitor network traffic and utilize threat intelligence to identify suspicious activities.	3.67	0.798
11	Banks collaborate with cybersecurity firms and share threat intelligence.	3.42	0.656
12	Banks maintain stringent physical security measures at their data centers and server locations.	3.48	0.683
13	Banks keep their systems up to date with the latest security patches and software updates.	3.31	0.990
14	Banks maintain a strong and properly configured firewall infrastructure to prevent unauthorized access to the network.	3.40	1.043
15	Banks utilize secure protocols, such as HTTPS and SSL/TLS, for online transactions and implement disk encryption.	3.47	0.927
16	Banks segment their network into multiple zones and restrict access between them to minimize the impact of unauthorized access.	3.41	1.044
17	Banks employ Data Loss Prevention (DLP) solutions to monitor and prevent the unauthorized transmission or storage of sensitive data.	3.52	0.817
Ensuring Network Security Infrastructure and Mitigating Risks of Electronic Crimes in Banks		3.40	0.442

It is noted from Table III that the arithmetic mean of the estimates of the sample members for how can banks effectively mitigate the risk of electronic crimes and ensure the security of customers' financial data. The results found about 81.5% of the sample agreed that Banks utilize robust encryption technologies to ensure the protection of customer financial data while it is being transmitted and stored with a mean of 3.70 and standard deviation of 0.806 and 80.7% of them ensure their banks implement systems that continuously monitor network traffic and utilize threat intelligence to identify suspicious activities with mean 3.67 and standard deviation 0.798. The results also showed that 68.1% of participants ensure banks employ Data Loss Prevention (DLP) solutions to monitor and prevent the unauthorized transmission or storage of sensitive data and 64.1% agree that their banks maintain a strong and properly configured firewall infrastructure to prevent unauthorized access to the network with mean 3.52, 3.40 and standard deviation 0.817, 1.043 respectively while 37.4% and 35.6% of respondent agree that Banks employ multi-factor authentication methods, such as hardware tokens and SMS-based verification, for customer logins and enforce strong password policies for both employees and customers.

The findings found a high agreement (agree) in 58.1% of the sample with a mean of 3.32 and standard deviation 1.136 which represent that biometric authentication techniques used in the banks like fingerprint and facial recognition to verify the identity of customers. On the other hand, the samples were asked about to which extent banks keep their systems up to date with the latest security patches and software updates and collaborate with cybersecurity firms and share threat intelligence, the results showed also a high agreement of 44% and 46.7% of the sample with a mean of 3.31 and 3.42 and standard deviation 0.990 and 0.656 accordingly. Regarding the second question of the study, which states "How strong is cyber security that can give banks a competitive advantage?" The arithmetic means and standard deviations of the sample's answers to the statements related to the competitive advantage of strong cybersecurity in banks were analyzed. Table IV summarizes the results.

The results in Table IV indicated the arithmetic mean of the estimates of the sample members for the degree of effects of strong cybersecurity on competitive advantage in banks ranged between (3.08) for paragraph No (6) and (3.77) for paragraph No (10). The results show that 83.3% of participants agree that banks with a strong cybersecurity system provide them with a secure platform for innovative digital banking services with a mean (of 3.77) and standard deviation of 0.752 and 66.3% of them believe banks with a strong cybersecurity system enhance the safeguarding of customer data and minimize the risk of data breaches with a mean (3.51) and standard deviation 0.895. Regarding reputation and loyalty, the results show 46.2% confirmed banks with a strong cybersecurity system can bolster reputation and foster customer loyalty with a mean (3.44) and a standard deviation of 0.791. Meanwhile, when the participants were asked if banks with a strong cybersecurity system are less susceptible to operational disruptions or downtime, the findings revealed a moderate agreement of

40.7% of the sample with a mean of 3.08 and a standard deviation of 0.921. Otherwise, 57.8% of the sample understand that Banks with a strong cybersecurity system mitigate financial losses resulting from cyber-attacks and 35.2% understand that banks with a strong cybersecurity system establish trust and confidence among bank-customers with 3.57 and 3.21 and standard deviation of 0.961 and 0.675. Accordingly, 50.3% of participants consider banks with a strong cybersecurity system to be better equipped to detect and respond to emerging threats with a mean of 3.10 and a standard deviation of 1.178.

TABLE IV. ARITHMETIC MEANS AND STANDARD DEVIATIONS ARE THE ESTIMATES OF THE STUDY SAMPLE ON THE COMPETITIVE ADVANTAGE OF STRONG CYBERSECURITY IN BANKS

No.	Statement	Mean	SD
1	Banks with a strong cybersecurity system establish trust and confidence among bank-customers.	3.21	0.675
2	Banks with a strong cybersecurity system enhance the safeguarding of customer data and minimize the risk of data breaches.	3.51	0.895
3	Banks with a strong cybersecurity system attract and retain customers.	3.63	0.839
4	Banks with a strong cybersecurity system mitigate financial losses resulting from cyberattacks.	3.57	0.961
5	Banks with a strong cybersecurity system comply with regulatory requirements and avoid penalties.	3.47	0.839
6	Banks with a strong cybersecurity system are less susceptible to operational disruptions or downtime.	3.08	0.921
7	Banks with a strong cybersecurity system set them apart from their competitors by fostering trust and dependability.	3.24	0.835
8	Banks with a strong cybersecurity system can bolster their reputation and foster customer loyalty.	3.44	0.791
9	Banks with a strong cybersecurity system are better equipped to detect and respond to emerging threats.	3.10	1.178
10	Banks with a strong cybersecurity system provide them with a secure platform for innovative digital banking services.	3.77	0.752
The Competitive Advantage of Strong Cybersecurity in Banks		3.39	0.446

Related to the third question of the study, which states "How can banks effectively educate and train employees and customers to be aware of and prevent cyber threats?" The arithmetic means and standard deviations of the sample's answers to the statements related to effective strategies for educating and training bank employees and customers were analyzed. Table V summarizes the results.

The results in Table V indicated the arithmetic mean of the estimates of the sample members for the degree of effects strategies for educating and training bank employees and customers ranged between (3.73) for paragraph No (2) and between (3.20) for paragraph No (5). The results show that 78.1% of participants agree that Banks can administer simulated phishing exercises to assess employees' awareness and response with a mean (3.73) and standard deviation of 0.541 and 71.1% of them believe banks can encourage the

reporting of suspicious activities or potential cyber threats with a mean (3.57) and standard deviation 1.010. Regarding regular updates, the results show 63.4% confirmed Banks can provide regular updates and reminders about emerging cyber threats and best practices with a mean (of 3.47) and a standard deviation of 0.923. Meanwhile, when the participants were asked if banks can offer frequent training sessions on cybersecurity to the employee, the findings revealed a moderate agreement of 36.3% of the sample with a mean of 3.24 and a standard deviation of 0.720. Otherwise, 48.9% of the sample understand that banks can collaborate with external cybersecurity experts to organize workshops and seminars for employees and customers and 50.8 % understand that banks can regularly assess the effectiveness of cybersecurity education and training programs with a mean of 3.20, 3.31, and standard deviation 1.078, 0.983. Accordingly, 54.8% of participants consider Banks can provide incentives or rewards for active participation in cybersecurity initiatives by employees and customers with a mean of 3.38 and a standard deviation of 0.920.

TABLE V. ARITHMETIC MEANS AND STANDARD DEVIATIONS ARE THE ESTIMATES OF THE STUDY SAMPLE ON EFFECTIVE STRATEGIES FOR EDUCATING AND TRAINING BANK EMPLOYEES AND CUSTOMERS

No.	Statement	Mean	SD
1	Banks can offer frequent training sessions on cybersecurity to employees.	3.24	0.720
2	Banks can administer simulated phishing exercises to assess employees' awareness and response.	3.73	0.541
3	Banks can provide regular updates and reminders about emerging cyber threats and best practices.	3.47	0.923
4	Banks can conduct cybersecurity awareness campaigns through email newsletters, social media, or other platforms.	3.38	0.817
5	Banks can collaborate with external cybersecurity experts to organize workshops and seminars for employees and customers.	3.20	1.078
6	Banks can encourage the reporting of suspicious activities or potential cyber threats.	3.57	1.010
7	Banks can provide incentives or rewards for active participation in cybersecurity initiatives by employees and customers.	3.38	0.920
8	Banks can regularly assess the effectiveness of cybersecurity education and training programs.	3.31	0.983
Effective Strategies for Educating and Training Bank Employees and Customers		3.41	0.509

To determine the importance of the role that cybersecurity plays in preventing and mitigating cybercrime, the ranking method was used to achieve this goal, and the importance was classified based on the relative importance index (RII). Table VI summaries the results.

TABLE VI. RANKING THE ROLE THAT CYBERSECURITY IN PREVENTING AND MITIGATING CYBERCRIME

Statements	RII	Ranking category	Overall ranking	Importance level
Banks regularly provide comprehensive training	0.608	17	35	M

programs to their employees on electronic crimes and protocols for cybersecurity.				
Banks educate their customers on the risks of electronic crimes, cybersecurity, and the best practices for safeguarding their financial information.	0.646	16	29	H-M
Banks employ multi-factor authentication methods, such as hardware tokens and SMS-based verification, for customer logins.	0.646	15	30	H-M
Banks embrace biometric authentication techniques like fingerprint and facial recognition to verify the identity of customers.	0.664	12	23	H-M
Banks utilize robust encryption technologies to ensure the protection of customer financial data while it is being transmitted and stored.	0.740	1	3	H-M
Banks enforce strong password policies for both employees and customers.	0.660	14	26	H-M
Banks conduct regular security audits and vulnerability assessments of their IT systems.	0.698	4	10	H-M
Banks employ real-time fraud detection and prevention systems.	0.686	7	16	H-M
Banks have well-defined incident response plans and robust systems in place to recover from electronic crimes.	.686	8	17	H-M
Banks implement systems that continuously monitor network traffic and utilize threat intelligence to identify suspicious activities.	.734	2	4	H-M
Banks collaborate with cybersecurity firms and share threat intelligence.	.684	9	18	H-M
Banks maintain stringent physical security measures at their data centers and server locations.	.696	5	11	H-M
Banks keep their systems up to date with the latest security patches and software updates.	.662	13	24	H-M
Banks maintain a strong and properly configured firewall infrastructure to prevent unauthorized access to the network.	.680	11	20	H-M

Banks utilize secure protocols, such as HTTPS and SSL/TLS, for online transactions and implement disk encryption.	0.694	6	12	H-M
Banks segment their network into multiple zones and restrict access between them to minimize the impact of unauthorized access.	.682	10	19	H-M
Banks employ Data Loss Prevention (DLP) solutions to monitor and prevent the unauthorized transmission or storage of sensitive data.	0.704	3	8	H-M

Based on the relative importance index, all methods related to cybersecurity methods in preventing and mitigating cybercrime were of medium to high importance, but the most important methods can banks effectively mitigate the risk of electronic crimes and ensure the security of customers' financial data that banks utilize robust encryption technologies to ensure the protection of customer financial data while it is being transmitted and stored (RII=0.740). Then banks implement systems that continuously monitor network traffic and utilize threat intelligence to identify suspicious activities (RII=0.734), banks employ Data Loss Prevention (DLP) solutions to monitor and prevent the unauthorized transmission or storage of sensitive data with (RII=0.704).

Educating customers about the dangers of cybercrime, cybersecurity, and best practices to protect their financial information was also ranked 16th in importance with (RII=0.646) and Banks regularly provide comprehensive training programs to their employees on electronic crimes and protocols for cybersecurity 17th in importance with (RII=0.608).

Related to how strong cyber security gives banks a competitive advantage, the ranking method was used to achieve this goal, and the importance was classified based on the relative importance index (RII). Table VII shows the results.

TABLE VII. RANKING THE ROLE OF CYBER SECURITY ON BANKS' COMPETITIVE ADVANTAGE

Statements	RII	Ranking by category	Overall ranking	Importance level
Banks with a strong cybersecurity system establish trust and confidence among bank-customers.	.642	8	31	H-M
Banks with a strong cybersecurity system enhance the safeguarding of customer data and minimizes the risk of data breaches.	0.702	4	9	H-M
Banks with a strong cybersecurity system attract and retain customers.	0.726	2	5	H-M

Banks with a strong cybersecurity system mitigate financial losses resulting from cyberattacks.	0.714	3	6	H-M
Banks with a strong cybersecurity system comply with regulatory requirements and avoid penalties.	0.694	5	13	H-M
Banks with a strong cybersecurity system are less susceptible to operational disruptions or downtime.	0.616	10	34	H-M
Banks with a strong cybersecurity system set them apart from their competitors by fostering trust and dependability.	0.648	7	27	H-M
Banks with a strong cybersecurity system can bolster their reputation and foster customer loyalty.	0.688	6	15	H-M
Banks with a strong cybersecurity system are better equipped to detect and respond to emerging threats.	0.620	9	33	H-M
Banks with a strong cybersecurity system provide them with a secure platform for innovative digital banking services.	0.754	1	1	H-M

It is clear from Table VII that banks that have a strong cyber security system provide a secure platform for digital financial services, which increases the competitive advantage as they were ranked first for their relative importance both at the category level and overall ranking with (RII=0.754). The results also showed the high importance of a strong cybersecurity system in attracting and retaining customers and mitigating financial losses resulting from cybercrime. With (RII=0.726) and (RII=0.714) accordingly, the results also showed that there is a high importance of a strong cyber security system in banks to protect customer data, reduce risks and data breaches with (RII=0.702), and enhance the bank's reputation and loyalty with (RII=0.688), detect and respond to emerging threats with (RII=0.620)

Regarding training and educating employees and customers to contribute effectively to preventing electronic crimes, the ranking method was used, and the importance was classified based on the relative importance index (RII). Table VIII shows the results.

TABLE VIII. RANKING THE ROLE OF TRAINING AND EDUCATING EMPLOYEES AND CUSTOMERS TO PREVENT ELECTRONIC CRIMES

Statements	RII	Ranking by category	Overall ranking	Importance level
Banks can offer frequent training sessions on cybersecurity to employees.	0.648	7	28	H-M
Banks can administer simulated phishing exercises to assess employees' awareness and response.	0.746	1	2	H-M
Banks can provide regular updates and reminders about emerging cyber threats and best practices.	0.694	3	14	H-M
Banks can conduct cybersecurity awareness campaigns through email newsletters, social media, or other platforms.	0.676	5	21	H-M
Banks can collaborate with external cybersecurity experts to organize workshops and seminars for employees and customers.	0.640	8	32	H-M
Banks can encourage the reporting of suspicious activities or potential cyber threats.	0.714	2	7	H-M
Banks can provide incentives or rewards for active participation in cybersecurity initiatives by employees and customers.	0.676	4	22	H-M
Banks can regularly assess the effectiveness of cybersecurity education and training programs.	0.662	6	25	H-M

The results showed the importance of training and educating employees and customers to contribute effectively to preventing electronic crimes. The statement that Banks can administer simulated phishing exercises to assess employees' awareness and response came in the highest rank of importance (RII=0.746). The results also showed high importance for banks to encourage the reports of suspicious activities or potential cyber threats with (RII=0.714). The results also showed that there is a high importance on providing regular updates and reminders about emerging

cyber threats and best practices with (RII= 0.694) providing incentives or rewards for active participation in cybersecurity initiatives by employees and customers with (RII=0.676), and conducting cybersecurity awareness campaigns through email newsletters, social media, or other platforms with (RII=0.676). Banks can offer frequent training sessions on cybersecurity to employees and assess the effectiveness of cybersecurity education and training programs also banks can collaborate with external cybersecurity experts to organize workshops and seminars for employees and customers all these methods have a highly important role in preventing electronic crimes.

V. DISCUSSION

The results showed that cybersecurity plays a crucial role in preventing and mitigating electronic crimes. Improving network security infrastructure contributes to preventing unauthorized access and data breaches like encryption technologies; employing data loss prevention (DLP) solutions to monitor and prevent the unauthorized transmission or storage of sensitive data; enforce strong password policies for both employees and customers, define incident response plans and robust systems to recover from electronic crimes; configured firewall infrastructure to prevent unauthorized access to the network; use biometric authentication techniques and employ real-time fraud detection and prevention systems. These results are consistent with study of Ghelani et al. [17], which recommended the establishment of a smart internet banking system and intruder detection through the use of biometric prints, fingerprints, passwords, OTPs, and other methods which reduces the number of threats.

The results also showed the importance of training and educating employees and customers to prevent electronic crimes banks can conduct cybersecurity awareness campaigns and banks can collaborate with external cybersecurity experts to organize workshops and seminars for employees and customers. These results agree with Sharma [49] who recommended the key to cyber security solutions is having well-trained staff and effective awareness campaigns to conduct a thorough analysis of risk management, an internal security task team should collaborate with a reliable security vendor. These study results also agreed with Al-Alawi & Al-Bassam's [50] study which shows mandating security awareness training is one of the most commonly used methods by boards of directors and executive managers to reduce cyber risks.

The finding shows that the knowledge and skills of the team of employees who deal with cyber attempts are important factors in determining the effectiveness of the cyber security method used. These results agreed with the study of Malik & Islam [51], which showed that cybercrime incidents harm organizational performance but awareness of information security reduces the negative impact of cybercrime and the roles of awareness related to information security in reducing cyber fraud and enhancing the security of customer information and the overall performance of financial institutions, also agreed with Khan [52] and Al-Daeef et al. [53] studies which emphasized the need to educate employees through programs aimed at educating employees about safe computing practices and the risks of human error that may

lead to security breaches; for example, phishing simulations that provide in-depth worker training scenarios are a proven method for increasing security awareness.

The results ensure that cybersecurity can boost customer trust by preventing sensitive customer data from being leaked and allowing a company to deliver on its promises, so if a bank had no breaches, it would be highly regarded and retained by its customers. These findings agreed with Kosutic [54] which recommended that companies could achieve strategic value that will be challenging to imitate by developing specific cybersecurity dynamic capabilities, and thereby achieve sustainable competitive advantage. Another study by Vijayalakshmi et al. [55] also confirmed this result.

The study also confirmed that one source of competitive advantage for a bank is to have cybersecurity to evaluate the current security measures and protect crucial data which is consistent with Kosutic [54] which explores the elements needed for cybersecurity implementation and management in an organization, as well as how cybersecurity can contribute to the competitive advantage of a company.

The study confirms that cybercrime is an emerging threat to IT facilities, the ever-changing nature of cybercrime and the associated development makes it increasingly difficult for policymakers and government institutions to implement cybercrime laws and policies, as a result, the banking industry must consistently educate its customers on information security techniques and how to avoid hacking into their accounts.

The study recommends to develop an alert system that can raise awareness for both banks and bank-customers whenever there is any possible entry or access to the customer's account or organization's confidential information. Furthermore, the banking industry must take into account the concept of effectively implementing and integrating big data technology into its system mitigating the negative consequences of cybercrime. This will allow for the storage of large files, and the data would aid in the examination, monitoring, and detection of network irregularities.

VI. CONCLUSION

The study's conclusions may directly affect how the Jordanian banking industry develops its regulatory and policy frameworks for electronic crime prevention and cyber security. The study's findings can help policymakers as they create recommendations and legislation. The report might provide banks in Jordan and possibly elsewhere with useful advice for developing effective risk management plans to protect against cybercrime. Based on the study's recommendations, banks can modify their cyber security procedures. The study's findings may help banks to be better equipped to defend against online attacks. Banks can discover weaknesses and put preventative measures in place to stop electronic crimes using the study's insights.

VII. LIMITATION

This study has several limitations; one of the most important is the amount of survey feedback received such as this study would have benefited from more responses. Despite

the limitation mentioned above the results obtained in this study is significant for Jordan's banking and financial institutions, which can use the findings to improve their employees' skills in detecting various cyber-attacks. Furthermore, these findings are critical in broadening the understanding of cybersecurity and its impact on financial matters for organizations. Also, the study was limited to cybersecurity with no examination of areas closely related to cybersecurity such as privacy, fraud, and physical security; these related areas may influence some cybersecurity models presented in this study.

REFERENCES

- [1] M. Lezzi and A. Corallo, "Cybersecurity for Industry 4.0 in the current literature: A reference framework," *Computers in Industry*, vol. 103, pp. 97–110, Dec. 2018, doi: 10.1016/j.compind.2018.09.004.
- [2] A. Ustundag, E. Cevikkan, B. C. Ervural, and B. Ervural, "Overview of cyber security in the industry 4.0 era," *Industry 4.0: managing the digital transformation*, pp. 267–284, 2018.
- [3] M. E. Gunduz and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer Networks*, vol. 169, p. 107094, Mar. 2020, doi: 10.1016/j.comnet.2019.107094.
- [4] R. S. Deora and D. Chudasama, "Brief study of cybercrime on an internet," *Journal of Communication Engineering & Systems*, vol. 11, no. 1, pp. 1–6, 2021.
- [5] S. A. Afaq, M. S. Husain, A. Bello, and H. Sadia, "A critical analysis of cyber threats and their global impact," in *Computational Intelligent Security in Wireless Communications*, Boca Raton: CRC Press, 2022, pp. 201–220.
- [6] J. Telo, "Understanding Security Awareness Among Bank Customers: A Study Using Multiple Regression Analysis," *Sage Science Review of Educational Technology*, vol. 6, no. 1, pp. 26–38, 2023.
- [7] M. C. Arcuri, M. Brogi, and G. Gandolfi, "How Does Cyber Crime Affect Firms? The Effect of Information Security Breaches on Stock Returns," in *ITASEC*, 2017, pp. 175–193.
- [8] D. Margiansyah, "Revisiting Indonesia's economic diplomacy in the age of disruption: Towards digital economy and innovation diplomacy," *J. ASEAN Stud.*, vol. 8, no. 1, p. 15, 2020.
- [9] M. Seete, "The digitisation of a firm process and its impact on corporate governance," *Indian Journal of Corporate Governance*, vol. 15, no. 2, pp. 280–294, 2022.
- [10] M. E. Bonfanti, "Artificial intelligence and the offence-defence balance in cyber security," in *Cyber Security: Socio-Technological Uncertainty and Political Fragmentation*. London: Routledge, 2022, pp. 64–79.
- [11] H. M. Alzoubi et al., "Cyber security threats on digital banking," in *2022 1st International Conference on AI in Cybersecurity (ICAIC)*, 2022.
- [12] V. Wang, H. Nnaji, and J. Jung, "Internet banking in Nigeria: Cyber security breaches, practices and capability," *Int. J. Law Crime Justice*, vol. 62, no. 100415, p. 100415, 2020.
- [13] T. M. Mbelli and B. Dwolatzky, "Cyber security, a threat to cyber banking in South Africa: An approach to network and application security," in *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)*, 2016.
- [14] A. L. Upashovna, "The Impact of Information Warfare on the Socio-Economic Development of Society and the Issue of Information Security," *European Journal of Innovation in Nonformal Education*, vol. 2, pp. 245–248, 2022.
- [15] N. M. Mallaboyev, Q. M. Sharifjanovna, Q. Muxammadjon, and C. Shukurullo, "Information Security Issues," in *Conference Zone*, pp. 241–245, 2022.
- [16] M. H. U. Sharif and M. A. Mohammed, "A literature review of financial losses statistics for cyber security and future trend," *World J. Adv. Res. Rev.*, vol. 15, no. 1, pp. 138–156, 2022.

- [17] D. Ghelani, T. K. Hua, and S. K. R. Koduru, "Cyber Security Threats, Vulnerabilities, and Security Solutions Models in Banking," *Authorea Preprints*, 2022.
- [18] I. Ashraf et al., "A survey on cyber security threats in IoT-enabled maritime industry," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2022.
- [19] B. Dash, M. F. Ansari, P. Sharma, and A. Ali, "Threats and opportunities with AI-based cyber security intrusion detection: A review," *Int. J. Softw. Eng. Appl.*, vol. 13, no. 5, pp. 13–21, 2022.
- [20] R. Montasari, "Cyber Threats and the Security Risks They Pose to National Security: An Assessment of Cybersecurity Policy in the United Kingdom. Countering Cyberterrorism: The Confluence of Artificial Intelligence," in *Cyber Forensics and Digital Policing in US and UK National Cybersecurity*, 2023, pp. 7–25.
- [21] H. S. Lallie et al., "Cyber security in the age of COVID-19: A timeline and analysis of cyber-crime and cyber-attacks during the pandemic," *Comput. Secur.*, vol. 105, no. 102248, p. 102248, 2021.
- [22] S. E. A. Ali, F.-W. Lai, R. Hassan, and M. K. Shad, "The Long-Run impact of information security breach announcements on investors' confidence: The context of efficient market hypothesis," *Sustainability*, vol. 13, no. 3, p. 1066, 2021.
- [23] M. R. Mphatheni and W. Maluleke, "Cybersecurity as a response to combating cybercrime: Demystifying the prevailing threats and offering recommendations to the African regions," *International Journal of Research in Business and Social Science*, vol. 11, no. 4, pp. 384–396, 2022.
- [24] A. A. Mughal, "Cybersecurity Architecture for the Cloud: Protecting Network in a Virtual Environment," *International Journal of Intelligent Automation and Computing*, vol. 4, no. 1, pp. 35–48, 2021.
- [25] F. Yan, Y. Jian-Wen, and C. Lin, "Computer network security and technology research," in *Seventh International Conference on Measuring Technology and Mechatronics Automation*, IEEE, 2015, pp. 293–296.
- [26] Z. El Mrabet, N. Kaabouch, H. El Ghazi, and H. Ghazi, "Cyber-security in smart grid: Survey and challenges," *Computers & Electrical Engineering*, vol. 67, pp. 469–482, 2018.
- [27] V. Burton-Howard, *Protecting small business information from cyber security criminals: A qualitative study, Doctoral dissertation*, Colorado Technical University, 2018.
- [28] M. Shohoud, "Study the effectiveness of ISO 27001 to mitigate the cyber security threats in the Egyptian downstream oil and gas industry," *J. Inf. Secur.*, vol. 14, no. 02, pp. 152–180, 2023.
- [29] S. Maesaroh, L. Kusumaningrum, N. Sintawana, D. P. Lazirkha, and R. Dinda, "Wireless network security design and analysis using Wireless Intrusion Detection System," *International Journal of Cyber and IT Service Management*, vol. 2, no. 1, pp. 30–39, 2022.
- [30] M. F. Nasution, "The Role of Civil Law in the Protection of Privacy and Personal Data," *Innovative: Journal of Social Science Research*, vol. 3, no. 2, pp. 3669–3679, 2023.
- [31] S. G. Prasad, M. K. Badrinayanan, and V. C. Sharmila, A Study on the adoption of Threat Prevention and Proactive Threat Monitoring Technologies for Securing the Information Technology Assets in India. 2023.
- [32] M. F. G. Rafea, A. Hamdan, R. Binsaddig, and E. Qasem, "The effects of cyber crime on E-banking," in *Digitalisation: Opportunities and Challenges for Business*, Cham: Springer International Publishing, 2023, pp. 560–569.
- [33] M. A. G. Zainal et al., "A decentralized autonomous personal data management system in banking sector," *Computers & Electrical Engineering*, vol. 100, p. 108027, May 2022, doi: 10.1016/j.compeleceng.2022.108027.
- [34] Singh, U. and Singh, P. (2022) 'Managing cyber security', *Journal of Management and Service Science (JMSS)*, 2(1), pp. 1–10. doi:10.54060/jmss/002.01.002.
- [35] V. Bandari, "Enterprise Data Security Measures: A Comparative Review of Effectiveness and Risks Across Different Industries and Organization Types," *International Journal of Business Intelligence and Big Data Analytics*, vol. 6, no. 1, pp. 1–11, 2023.
- [36] J. Wolff, "Trends in cybercrime during the COVID-19 pandemic," in *Beyond the Pandemic? Exploring the Impact of COVID-19 on Telecommunications and the Internet*, Emerald Publishing Limited, 2023, pp. 215–227.
- [37] B. Ritchot, "An enterprise security program and architecture to support business drivers," *Technol. Innov. Manag. Rev.*, vol. 3, no. 8, pp. 25–33, 2013.
- [38] M. Qasaimeh, R. A. Hammour, M. B. Yassein, R. S. Al-Qassas, J. A. L. Torralbo, and D. Lizcano, "Advanced security testing using a cyber-attack forecasting model: A case study of financial institutions," *J. Softw. (Malden)*, vol. 34, no. 11, 2022.
- [39] Srivastava, G., Jhaveri, R. H., Bhattacharya, S., Pandya, S., Maddikunta, P. K. R., Yenduri, G., ... & Gadekallu, T. R. (2022). XAI for cybersecurity: state of the art, challenges, open issues and future directions. arXiv preprint arXiv:2206.03585.
- [40] Y. Rivaldo, S. V. Kamanda, and E. Yusman, "The Influence Of Brand Image, Promotion And Trust On Customer Loyalty At Bank BSI Nagoya Batam Branch," *Jurnal Mantik*, vol. 6, no. 2, pp. 2385–2392, 2022.
- [41] A. Shukla, B. Katt, L. O. Nweke, P. K. Yeng, and G. K. Weldehawaryat, "System security assurance: A systematic literature review," *Comput. Sci. Rev.*, vol. 45, no. 100496, p. 100496, 2022.
- [42] M. Bidgoli, B. P. Knijnenburg, J. Grossklags, and B. Wardman, "Report now. Report effectively. Conceptualizing the industry practice for cybercrime reporting," in *2019 APWG Symposium on Electronic Crime Research (eCrime)*, 2019.
- [43] F. E. Catota, M. G. Morgan, and D. C. Sicker, "Cybersecurity incident response capabilities in the Ecuadorian financial sector," *Journal of Cybersecurity*, vol. 4, no. 1, 2018.
- [44] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing attacks: A recent comprehensive study and a new anatomy," *Front. Comput. Sci.*, vol. 3, 2021.
- [45] M. Alohali, N. Clarke, F. Li, and S. Furnell, "Identifying and predicting the factors affecting end-users' risk-taking behavior," *Information & Computer Security*, vol. 26, no. 3, pp. 306–326, 2018.
- [46] N. C. Roy and S. Prabhakaran, "Sustainable response system building against insider-led cyber frauds in banking sector: A machine learning approach," *Journal of Financial Crime*, 2022.
- [47] A. C. de Souza, N. M. C. Alexandre, and E. de B. Guirardello, "Psychometric properties in instruments evaluation of reliability and validity," *Epidemiologia e servicos de saude*, vol. 26, no. 3, pp. 649–659, 2017.
- [48] J. J. Vaske, J. Beaman, and C. C. Sponarski, "Rethinking internal consistency in cronbach's alpha," *Leisure sciences*, vol. 39, no. 2, pp. 163–173, 2017.
- [49] A. Sharma and P. Tandekar, "Cyber Security and Business Growth," in *Advances in Business Information Systems and Analytics*, IGI Global, 2016, pp. 14–27.
- [50] A. I. Al-Alawi and M. S. A. Bassam, "The significance of cybersecurity system in helping managing risk in banking and financial sector," *Journal of Xidian University*, vol. 14, no. 7, pp. 1523–1536, 2020.
- [51] Malik, M.S. and Islam, U. (2019), "Cybercrime: an emerging threat to the banking sector of Pakistan", *Journal of Financial Crime*, Vol. 26 No. 1, pp. 50-60.
- [52] M. J. Khan, "Securing network infrastructure with cyber security," *World J. Adv. Res. Rev.*, vol. 17, no. 2, pp. 803–813, 2023.
- [53] M. M. Al-Daeef, N. Basir, and M. Saudi, "Security awareness training: A review," *Proceedings of the World Congress on Engineering*, vol. 1, pp. 5–7, 2017.
- [54] D. Kosutic and F. Pigni, "Cybersecurity: investing for competitive outcomes," *J. Bus. Strategy*, vol. 43, no. 1, pp. 28–36, 2022.
- [55] P. Vijayalakshmi and D. Karthika, "A COMPARATIVE STUDY ON CYBER SECURITY THREATS DETECTION IN INTERNET OF THINGS," *A COMPARATIVE STUDY ON CYBER SECURITY THREATS DETECTION IN INTERNET OF THINGS. ICTACT Journal on Communication Technology*, vol. 12, no. 2, 2021.

Research on the Local Path Planning for Mobile Robots Based on PRO-Dueling Deep Q-Network (DQN) Algorithm

Yaoyu Zhang, Caihong Li*, Guosheng Zhang, Ruihong Zhou, and Zhenying Liang
School of Computer Science and Technology, Shandong University of Technology, Zibo 255049, China

Abstract—This paper proposes a Pro-Dueling DQN algorithm to solve the problems of slow convergence speed and waste of effective experience of the traditional DQN (Deep Q-Network) algorithm for the local path planning of mobile robot. The new algorithm introduces a priority experience playback mechanism based on SumTree to avoid forgetting the learning effective experiences as the number of samples in the experience pool increases. A more detailed reward and punishment function is designed for the new algorithm to reduce the blindness of extracting experience in the early stages of algorithm training. The feasibility of the algorithm is verified by comparative verification on ROS simulation platform and real scene, respectively. The results show that the designed Pro-Dueling DQN algorithm converges faster and the length of planned path is shorter than that of the original DQN algorithm.

Keywords—Deep Q-Network (DQN) algorithm; local path planning; mobile robot; Pro-Dueling DQN algorithm; SumTree

I. INTRODUCTION

It is crucial for robots to avoid obstacles and plan effective paths in the research of mobile robot navigation. There are many effective path planning methods for obstacle avoidance at present. The traditional methods mainly include A* algorithm [1], Dijkstra algorithm [2], fuzzy control algorithm [3], genetic algorithm [4], artificial potential field method [5] and neural network [6]. Reinforcement learning algorithm [7] has received widespread attention because it can solve the shortcomings of traditional algorithms such as strong dependence on environment in robot path planning. It does not require any prior knowledge, and optimizes the strategy by interacting with the environment and accumulating rewards. The combination of deep learning [8] and reinforcement learning has extended traditional reinforcement learning to multidimensional state space and action space in recent years. Deep reinforcement learning [9] combines the ability of deep learning algorithm to understand perception problems and the ability to fit the learning results of reinforcement learning algorithm [10]. It has been widely used in robot path planning research.

Q-learning algorithm is one of the reinforcement learning algorithms proposed by Watkins, which is independent of environmental prior model in the path planning problem of mobile robot[11]. However, the strategy of storing the state-action value function by a Q-value table will cause the disaster of dimension as the environment states become more and more complex. Mnih et al. proposed Deep Q-Network (DQN),

which combined CNN (Convolutional Neural Networks) with Q-learning algorithm to solve the dimension disaster of Q-learning method, and pushed the research of deep reinforcement learning to a new level [12]. Z. Wang et al. innovated the network structure on the basis of DQN and divided the network into two parts: value function and advantage function to reduce the excessive dependence of states on the environment [13]. J.F. Zheng et al. proposed an improved DQN algorithm based on depth image information. PTZ (Pan/Tilt/Zoom) was used to obtain depth image information of obstacles, which improved the convergence speed of the network, but the stability and computational speed of the algorithm could not be guaranteed [14]. Xiaofei Yang et al. proposed a global path planning algorithm based on DDQN, which integrated an action mask method to deal with the invalid actions generated by the amphibious unmanned vehicle, but the algorithm training speed is not ideal[15]. Meng Guan et al. proposed a DQN path planning method combining heuristic reward and adaptive exploration strategy, designed a heuristic reward function based on artificial potential field method, and self-adaptively adjusted the balance between exploration and utilization in the algorithm, which accelerated the learning efficiency of the algorithm. However, the algorithm verification only stayed in the simulation stage, and the efficiency of the algorithm has not been verified in the real scene [16]. The new method improves the efficiency of the algorithm's exploration, but the length of exploration in the direction of exploration increases, resulting in excessive spatial dimensions.

This paper presents a Pro-Dueling DQN algorithm to solve the problems of poor convergence and waste of effective experience in the local path planning of mobile robot using DQN method. The research modifies the DQN neural network structure to combine the state value and action value to obtain a more accurate Q-value. The priority experience playback strategy [17] based on SumTree is adopted to give priority to the samples in the experience pool, and designs a reward and punishment function to solve the convergence difficulty problem caused by sparse rewards in unknown environments. This improves the utilization rate of effective experience in the algorithm, avoids the problem of local optimal solution, and accelerates the convergence speed of the algorithm. Comparing the convergence speed and planned path length of the two algorithms in simulation and real environments, experimental results show that the Pro-Dueling DQN algorithm performs better in various scenarios.

II. DQN ALGORITHM

DQN algorithm combines Q-learning algorithm with deep learning, uses network structure in deep learning to predict Q-value, and generates Q-table dynamically. It not only avoids the disaster of dimensionality in complex space, but also solves the instability problem of approximate representation of value functions for nonlinear functions to a certain extent [18]. Fig. 1 shows the process of DQN algorithm.

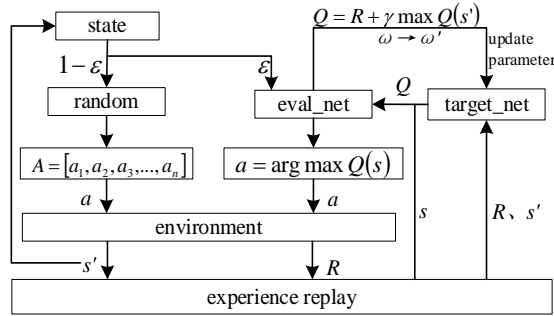


Fig. 1. DQN algorithm process diagram.

The algorithm defines two relatively independent networks with the same structure, namely *eval_net* and *target_net*. The agent interacts with the environment to achieve learning of the training network. All the parameters in the training network are assigned to the target network after the training of a fixed number of steps. The algorithm sets an experience replay unit to reduce the correlation of training samples and improve the instability of the action value function of neural network approximation reinforcement learning [19]. A batch of samples are evenly selected from the experience library and mixed together with the training samples to break the correlation between adjacent training samples and improve the utilization rate of the samples during each training. Where, the LOSS function is:

$$LOSS = \frac{1}{2} (Q_{\omega}(s) - (R + \gamma \max_{\omega'} Q_{\omega'}(s')))^2 \quad (1)$$

In Eq. (1), ω is a parameter in the *eval_net*, ω' is a parameter in the *target_net*. The parameters of *target_net* are synchronized with the training network every N steps to make the updated target more stable, that is, $\omega' \leftarrow \omega$.

III. PRO-DUELING DQN ALGORITHM

This research proposes a Pro-Dueling DQN algorithm to improve the training speed and convergence of DQN algorithm. Two different branches are introduced at the back end of DQN neural network to predict the value of state and action, then the results of these two branches are combined to output the Q-value to reduce the dependence of action on state. In addition, the priority experience playback based on SumTree replaces the uniform sampling playback mechanism of DQN algorithm to increase the sampling rate of important samples. The Pro-Dueling DQN algorithm includes the design of network structure of state space and action space, reward and punishment function and priority experience playback.

A. The Design of State Space and Action Space

The state space is the feedback of the environment information of the mobile robot. The input of the network is the state vector. The robot selects the subsequent action based on the state information, obtains the corresponding reward or punishment, and optimizes the strategy by accumulating the reward value.

The laser radar installed on the robot detects the surrounding environmental information. The detection range of radar sensor is 180°, and a group of data is returned every 15° with a total of 12 groups. Fig. 2 shows the laser radar information.

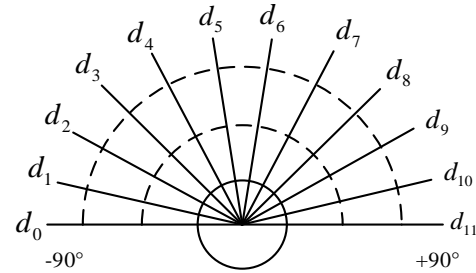


Fig. 2. Laser radar information.

The position information of the robot consists of the obstacle distance information returned by the radar in 12 directions d_n ($n=0\sim 11$), the distance between the robot and the target point D_g , and the angle between the robot and the target point θ_g . The state space of the robot is defined as:

$$S = (d_n, D_g, \theta_g) \quad (2)$$

The robot's action space includes action information in five directions, defined as:

$$A = \{a_t, t = 0 \sim 4\} \quad (3)$$

The linear speed of the robot is constant, set as 0.15m/s, and the angular velocity is determined by its action. The relationship between the robot angular velocity (Angle_v) corresponding to the five action values of the robot is shown in Table I.

TABLE I. CORRESPONDING RELATIONSHIP BETWEEN ROBOT ACTION AND ROTATION ANGLE

Action	Angle_v (rad/s)
0	-1.5
1	-0.75
2	0
3	0.75
4	1.5

B. Network Structure

The neural network used by Pro-Dueling DQN algorithm contains 14 inputs of the state space and 5 outputs of the action space, as shown in Fig. 3.

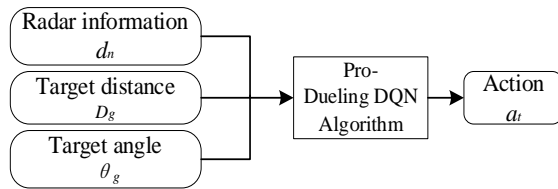


Fig. 3. Input and output of Pro-Dueling DQN network.

In the Pro-Dueling DQN algorithm, the output of the network includes a value function and an advantage function. The formula is as follows:

$$Q(s, a; \gamma, \alpha, \beta) = V(s; \gamma, \beta) + A(s, a; \gamma, \alpha) \quad (4)$$

In Eq. (4), $V(s; \gamma, \beta)$ is a value function, $A(s, a; \gamma, \alpha)$ is the advantage function of taking different actions in this state, indicating the difference of taking different actions. γ is a network structure, α is the parameter of value function, β is the parameter of advantage function. It can be seen from the formula that $V(s; \gamma, \beta)$ function is only related to the state, and $A(s, a; \gamma, \alpha)$ depends on both state and action. The neural network outputs the value function and dominance function, respectively, and sums them to obtain the Q value. The robot only pays attention to the value of the state in some cases, and does not care about the difference caused by different actions by modeling $V(s; \gamma, \beta)$ function and $A(s, a; \gamma, \alpha)$ function. The approach works better with states that are less associated with an action.

Fig. 4 shows the network structure of the Pro-Dueling DQN algorithm. L_1 and L_2 are fully connected layers, which contains 128 and 64 hidden neuron nodes, respectively. In the network, input eigenvalues are used to obtain eigenvectors using a convolutional network. When outputting, two fully connected layers are used to correspond to the state value and advantage value, respectively. Finally, the state value and advantage value are added to obtain the action value of each action.

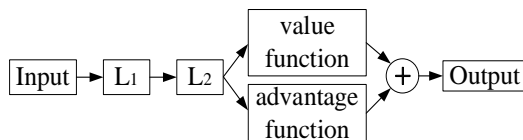


Fig. 4. Pro-Dueling DQN network structure.

C. The Design of Reward and Punishment Function

In the process of reinforcement learning, rewards and punishments obtained by mobile robot in its interactions with the environment are the key to completing tasks. In this research, the reward and punishment function is further refined, which including two parts: R_θ , the reward and punishment function of the angle between the robot and the target point, and R_t , the reward and punishment function of the distance between the robot and the obstacle. The final reward and punishment value R is obtained as follows by adding the two parts:

$$R_\theta = \begin{cases} C, & -\frac{1}{2}\pi < \theta < \frac{1}{2}\pi \\ -C, & \text{else} \end{cases} \quad (5)$$

$$R_t = \begin{cases} r_{goal}, & d_c < c_d \\ r_{collide}, & \min_x < c_o \end{cases}$$

$$R = R_\theta + R_t$$

In Eq. (5), C is a positive integer, θ is the angle of the robot to the target point, r_{goal} is a positive integer which representing a positive reward of the robot reaching the target point, d_c is the actual distance from the robot to the target, c_d is the threshold of reaching the target point. It means that the robot has reached the target point, when d_c is less than c_d . $r_{collide}$ is a negative integer, representing the penalty for the robot to encounter with obstacles, \min_x is the radar minimum, c_o is the safe distance. It is determined that the robot collides with obstacle as the radar value is less than the safe distance.

D. Priority Experience Playback Based on SumTree

The research adopts the priority experience replay strategy to improve this situation. The most valuable experiences are extracted first as training. TD -error determines the priority of the sample. The target function is weighted according to the TD -error of the sample. The greater the deviation, the larger the sample weight, and the higher the priority p is.

A random sampling method combined greedy sampling and uniformly distributed table sampling is used to solve the overfitting problem caused by greedy priority in the process of function approximation. This method ensures that the probability of sampling from the storage container is monotonous, and the lowest priority sample has a non-zero probability of being drawn. The sampling probability is as follows:

$$P(i) = \sum_k^{P_i^\alpha} P_k^\alpha \quad (6)$$

In Eq. (6), P_i is the priority of the i th sample, P_k is the priority of any sample, α is used to adjust the degree of priority. It is reduced to uniform sampling, when $\alpha = 0$. k is the number of batches of samples.

IV. ANALYSIS AND VERIFICATION OF PRO-DUELING DQN ALGORITHM

In this research, the environments of discrete obstacles, U-shaped obstacle and mixed obstacles are set for training to verify the feasibility of the designed Pro-Dueling DQN algorithm. The algorithm is compared in different environments with the traditional DQN algorithm. The environments are built on Gazebo of ROS platform. Their informations are projected into Rviz. The path planned by the mobile robot from the starting point to the target point is displayed on Rviz. In Rviz, the initial position of the robot is the starting point, the gray box represents the target point, the gray cylinder shows the obstacle, and the blue solid line

expresses the trajectory of the robot. A comparison graph of the average return value of each round of the two algorithms is drawn to observe the convergence of the algorithm more clearly and intuitively. The horizontal axis represents the number of training, and the vertical axis expresses the average reward of each training. At the same time, the path lengths of the robot by the two algorithms for path planning are recorded. Table II illustrates the parameter settings in the experiment.

TABLE II. EXPERIMENTAL PARAMETER SETTING

Parameter	Initialization value
learning rate	0.0001
attenuation factor	0.999
experience pool capacity	10000
number of learning experiences per round	128
maximum number of steps per round	400

A. Simulation Verification in Discrete Obstacles Environment

Fig. 5 shows the 7m×7m discrete obstacles environment set in Gazebo. The initial position of the robot is (-2.5, -2.5), and the target point coordinate is (2.2).

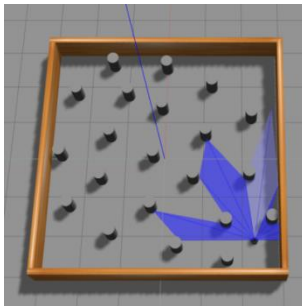


Fig. 5. The discrete obstacles simulation environment.

In the discrete obstacles environment, the two algorithms conduct 200 rounds of training, and record the average reward of each round. Fig. 6 shows the recording results. It can be seen from the figure that the average reward of the Pro-Dueling DQN algorithm gradually increases after 25 rounds, indicating that the success rate of the robot in the process of finding the target point is getting higher and higher, and the algorithm tends to converge after the 100th round of training. The convergence speed of it is faster than the traditional DQN algorithm, and the average reward of the convergence algorithm has less fluctuation and is relatively stable.

Fig. 7 shows the paths planned by the model after convergence of the two algorithms. The path planned by the Pro-Dueling DQN algorithm from the starting point to the target point is smoother and has fewer path steps than the traditional one, as can be seen in the figure. The number of steps taken by the designed strategy is 235, while the number is 283 by the traditional one.

B. Simulation Verification in U-shaped Obstacle Environment

Fig. 8 shows a 5m×5m U-shaped obstacle environment set up in Gazebo. The starting point coordinate of the robot is (-2.0, 0.0) and the target point coordinate is (1.5, 1.5).

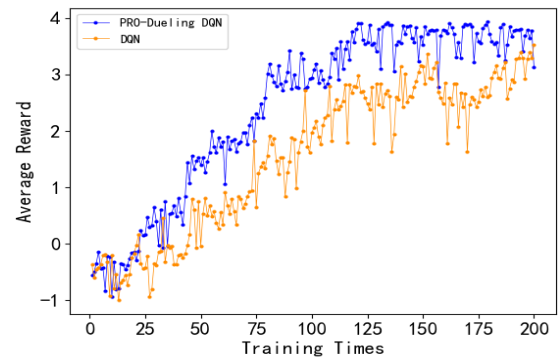
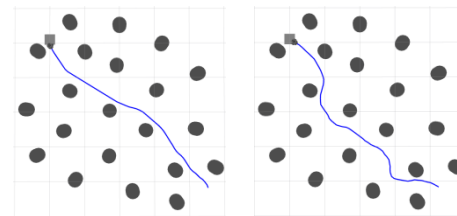


Fig. 6. The comparison of average reward per round in discrete obstacles environment.



(a) Pro-Dueling DQN (b) DQN

Fig. 7. Path planning in discrete obstacles environment.

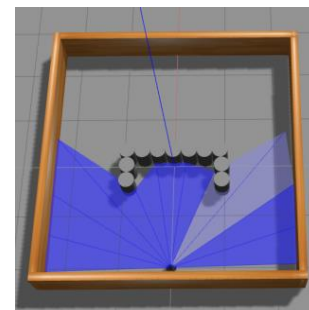


Fig. 8. U-shaped obstacle simulation environment.

Fig. 9 shows the record of the average return value of each round after 400 rounds of training by the Pro-Dueling DQN algorithm and the traditional DQN algorithm. Data displayed in the figure show that the average reward of the Pro-Dueling DQN algorithm after the 50th round is significantly higher than that of the traditional DQN algorithm, which indicates that the robot by the Pro-Dueling DQN algorithm can reach the target point more times. The Pro-Dueling DQN algorithm tends to converge after 150 rounds. The convergence speed of the Pro-Dueling DQN algorithm is faster, and the average reward of the convergence algorithm fluctuates less, indicating that the designed algorithm is more stable.

Fig. 10 shows the paths planned by the model after convergence of the two algorithms. The path planned by the Pro-Dueling DQN algorithm from the starting point to the target point is shorter than that by the traditional DQN algorithm. The number of steps is 268, while the number of the traditional one is 298.

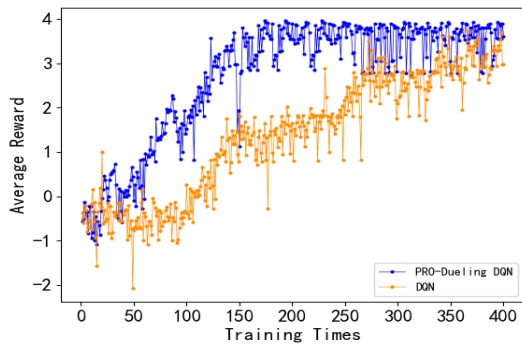


Fig. 9. Comparison of average reward per round in U-shaped obstacle environment.

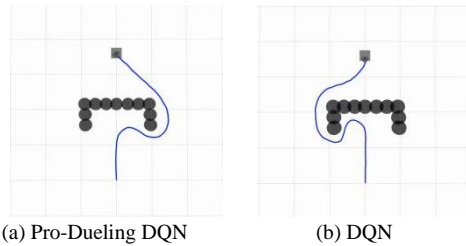


Fig. 10. U-shaped obstacle path planning.

C. Simulation Verification in Mixed Obstacles Environment

Fig. 11 and Fig. 12 show two 6m×6m mixed obstacles environments set up in Gazebo, which include discrete obstacles, 1-shaped obstacles and U-shaped obstacles. In Fig. 11, the starting point coordinate of the robot is (-2.0, 2.0) and the target point coordinate is (2.1,-2.0). In Fig. 12, the starting point coordinate of the robot is (-2.0,-2.0) and the target point coordinate is (2.5, 2.0).

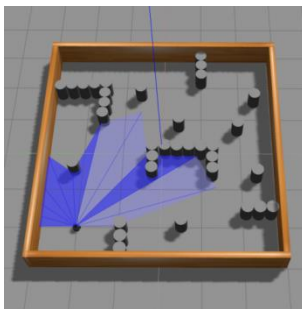


Fig. 11. Mixed obstacles environment (1).

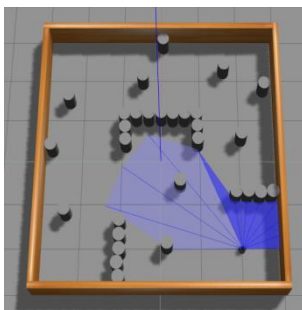


Fig. 12. Mixed obstacles environment (2).

Fig. 13 records the average reward of 500 rounds of training in the environments conducted by the Pro-Dueling DQN algorithm and the traditional DQN algorithm. The data in the figure show that the average reward value of the Pro-Dueling DQN algorithm after 100th round is significantly higher than that of the traditional one, indicating that the robot reach the target point more times by the Pro-Dueling DQN algorithm. The Pro-Dueling DQN algorithm tends to converge after 380 rounds. The convergence speed of it is faster than the traditional one.

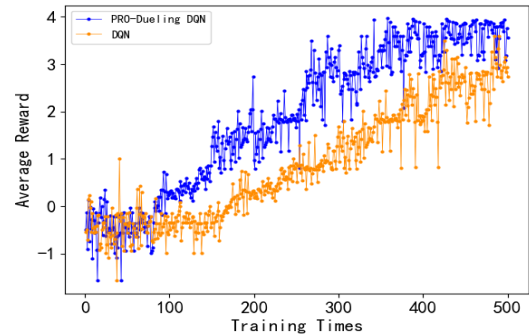


Fig. 13. Comparison of average reward value per round in mixed obstacles environment.

Fig. 14 and Fig. 15 show the path planned by the model after convergence of the two algorithms. The paths planned by the Pro-Dueling DQN algorithm from the starting point to the target point are shorter than the paths planned by the traditional DQN algorithm in both figures. In Fig. 14, the mixed obstacles environment (1), the number of steps taken by the Pro-Dueling DQN algorithm is 258, while the number by the traditional DQN algorithm is 311. The data in Fig. 15, the other mixed obstacles environment, are 302 and 337, respectively.

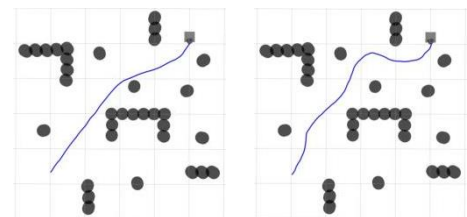


Fig. 14. Path planning in mixed obstacles environment (1).

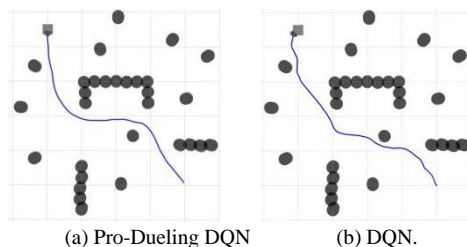


Fig. 15. Path planning in mixed obstacles environment (2).

The feasibility and effectiveness of the designed Pro-Dueling DQN algorithm in robot path planning are verified through simulation training in different obstacle environments, and are compared with the traditional DQN algorithm. The

simulations show that the convergence speed of Pro-Dueling DQN algorithm is faster and more stable than DQN algorithm. The path planned by the Pro-Dueling DQN algorithm is shorter and smoother in the same training times and operating environment.

Fig. 16 shows the comparison of the planned steps of the two algorithms from the starting point to the target point in the above three simulation environments. The data in the figure show that the Pro-Dueling DQN algorithm uses fewer steps than the traditional DQN algorithm in each environment.

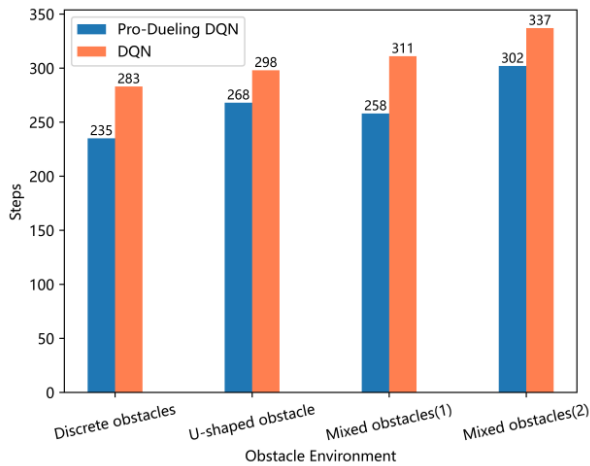


Fig. 16. Comparison of path steps between Pro-Dueling DQN algorithm and DQN algorithm.

D. Verification Experiments in Real Scenarios

The trained algorithm model is loaded into the robot, and the paths planned by the Pro-Dueling DQN algorithm and the DQN algorithm are tested and compared in the real environment. The ROS integrated SLAM (Simultaneous Localization and Mapping) function package is used to build the corridor environment model of teaching building. Fig. 17 shows the corridor environment and map model.

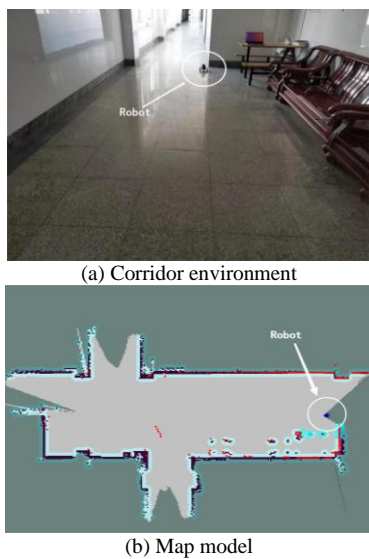


Fig. 17. Corridor environment.

The path planning experiments are carried out in the map model built by the robot. Fig. 18 shows the addition of temporary obstacles in the corridor environment during the experiment to verify the real-time obstacle avoidance performance of the local path planning algorithm. The robot uses Radar to scan obstacle information, and the pink parts in the map model represent the unknown obstacles detected in real time. The target point is selected in the map, and the robot moves towards the target point from the starting point. The obstacle information is detected and fed back in real time by Radar, so that the robot plans a collision-free path from the starting point to the target point. Rviz is used to display and record the planned paths. In the environment, the path length planned by the Pro-Dueling DQN algorithm is 7.835 meters, and that by the DQN algorithm is 8.563 meters. Both paths planned by the two algorithms can avoid temporary obstacles to reach the target point, while the paths planned by the Pro-Dueling DQN algorithm are shorter than those planned by the DQN algorithm, and the obstacle avoidance paths are smoother when encountering obstacles.



Fig. 18. Obstacle environment.

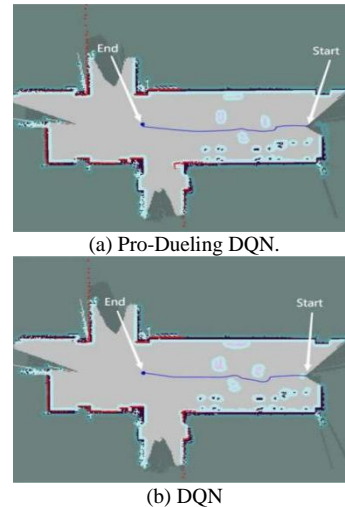


Fig. 19. Results of path planning in corridor environment.

V. CONCLUSION

This paper proposes a Pro-Dueling DQN algorithm based on DQN algorithm and SumTree algorithm to solve the local path planning problem of mobile robot in unknown environment. The effectiveness of the proposed algorithm is verified by comparison experiments on ROS simulation platform and real environment. The experimental results show that the trained Pro Dueling DQN algorithm model can perform better in robot local path planning tasks, obtain

smoother paths compared to the original algorithm, and complete tasks more efficiently and quickly. At the same time, the model has certain adaptability and can plan feasible paths in unknown environments in real-time based on sensor information.

However, the robot designed in this research has fewer actions, resulting in a large swing range in the training process. The planned paths in the complex and dense obstacles environments are not smooth, and the trained paths are not the optimal shortest ones. So future work will design more detailed action spaces, increase training time, and enable robot to plan better paths.

ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of Shandong Province, China (No. ZR2021MF072).

REFERENCES

- [1] T.T. Sang, J.C. Xiao, J.F. Xiong, H.Y. Xia, and Z.Z. Wang, Path planning method of unmanned surface vehicles formation based on improved A* algorithm, *Journal of Marine Science and Engineering*, 2023, vol. 11, no. 1, pp. 176-176.
- [2] B.Y. He, Application of Dijkstra algorithm in finding the shortest path, *Journal of Physics: Conference Series*, 2022, vol. 2181, no. 1.
- [3] H.B. Gao, S.Y. Lu, and T. Wang, Motion path planning of 6-DOF industrial robot based on fuzzy control algorithm, *Journal of Intelligent & Fuzzy Systems*, 2020, vol. 38, no. 4, pp. 3773-3782.
- [4] K. Hao, J.L. Zhao, Z.S. Li, Y.L. Liu, and L. Zhao, Dynamic path planning of a three-dimensional underwater AUV based on an adaptive genetic algorithm, *Ocean Engineering*, 2022, vol. 263.
- [5] G.Q. Zhang, J. Han, J.Q. Li, and X.K. Zhang, APF-based intelligent navigation approach for USV in presence of mixed potential directions: Guidance and control design, *Ocean Engineering*, 2022, vol. 260.
- [6] W.Z. Du, Q.M. Zhang, Z.X. He, and X. Wang, Real Time Neural Network Path Planning Algorithm for Robot, *International Journal of Frontiers in Engineering Technology*, 2021, vol. 3, no. 5.
- [7] A. Khan, F. Jiang, S. Liu, and O. Ibrahim, Playing a FPS doom video game with deep visual reinforcement learning, *Automatic Control and Computer Sciences*, 2019, vol. 53, no. 3, pp. 214-222.
- [8] B. Tamir, J. William, and Y. Kazuya, Deep learned path planning via randomized Reward-Linked-Goals and potential space applications, *CoRR*, 2019, vol. abs/1909.06034.
- [9] Z. Li, S.H. Yuan, X.F. Yin, X.Y. Li, and S.X. Tang, Research into autonomous vehicles following and obstacle avoidance based on deep reinforcement learning method under map constraints, *Sensors*, 2023, vol. 23, no. 2, pp. 844-844.
- [10] Z. Yu, J. Bi, and H.T. Yuan, A path planning method for complex naval battlefields based on improved DQN algorithm, *Journal of Intelligent Science and Technology*, 2022, vol. 4, no. 3, pp. 418-425.
- [11] C. Watkins, and P. Dayan, Q-learning, *Machine Learning*, 1992, vol. 8, no. 3-4, pp. 279-292.
- [12] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Grabes, M. Riedmiller, A.K. Fidjeland, and G. Ostrovski, Human-level control through deep reinforcement learning, *Nature*, 2015, vol. 518, no. 7540, pp. 529-533.
- [13] Z. Wang, N.D. Freitas, and M. Lanctot, Dueling network architectures for deep reinforcement learning, *CoRR*, 2015, vol. abs/1511.06581.
- [14] J.F. Zheng, S.R. Mao, Z.Y. Wu, P.C. Kong, and H. Qiang, Improved path planning for indoor patrol robot based on deep reinforcement learning, *Symmetry*, 2022, vol. 14, no. 1, pp. 132-132.
- [15] X.F. Yang, Y.L. Shi, W. Liu, H. Ye, W.B. Zhong, and Z.R. Xiang, Global path planning algorithm based on double DQN for multi-tasks amphibious unmanned surface vehicle, *Ocean Engineering*, 2022, vol. 266, no. P1.
- [16] M. Guan, F.X. Yang, J.C. Jiao, and X.P. Chen, Research on path planning of mobile robot based on improved Deep Q Network, *Journal of Physics: Conference Series*, 2021, vol. 1820, no. 1, pp. 012024-.
- [17] Y.Y. Zhang, X.P. Rao, C.Y. Liu, X.B. Zhang, and Y. Zhou, A cooperative EV charging scheduling strategy based on double deep Q-network and Prioritized experience replay, *Engineering Applications of Artificial Intelligence*, 2023, vol. 118.
- [18] E. Erkan, and M.A. Arserim, Mobile robot application with hierarchical start position DQN, *Computational Intelligence and Neuroscience*, 2022, vol. 2022.
- [19] Y.B. Chen, D.C. Li, H.G. Zhong, O.W. Zhu, and Z.Q. Zhao, The determination of reward function in AGV motion control based on DQN, *Journal of Physics: Conference Series*, 2022, vol. 2320, no. 1.

Prostate Cancer Detection and Analysis using Advanced Machine Learning

Mowafaq Salem Alzboon, Mohammad Subhi Al-Batah

Faculty of Science and Information Technology, Jadara University, Irbid, Jordan

Abstract—Prostate cancer is one of the leading causes of cancer-related deaths among men. Early detection of prostate cancer is essential in improving the survival rate of patients. This study aimed to develop a machine-learning model for detecting and diagnosing prostate cancer using clinical and radiological data. The dataset consists of 200 patients with prostate cancer and 200 healthy controls and extracted features from their clinical and radiological data. Then, the data trained and evaluated using several machine learning models, including logistic Regression, decision tree, random forest, support vector machine, and neural network models, using 10-fold cross-validation. Our results show that the random forest model achieved the highest accuracy of 0.92, with a sensitivity of 0.95 and a specificity of 0.89. The decision tree model achieved a nearly similar accuracy of 0.91, while the logistic regression, support vector machine, and neural network models achieved lower accuracies of 0.86, 0.87, and 0.88, respectively. Our findings suggest that machine learning models can effectively detect and diagnose prostate cancer using clinical and radiological data. The random forest model may be the most suitable model for this task.

Keywords—Prostate cancer; machine learning; clinical data; radiological data; diagnosis; medical diagnosis

I. INTRODUCTION

The use of machine learning techniques to study cancer has produced encouraging results, with the promise of more precise and time-saving approaches for identifying malignant cells and forecasting patient outcomes [1]. Machine learning algorithms can analyze data in search of patterns and attributes characteristic of malignant cells or tumours and then indicate how likely the disease will spread or reoccur [2, 3, 4]. Machine learning algorithms can identify small changes in tissue structure that may suggest the presence of malignant cells. They can forecast the likelihood of cancer progression and offer viable treatment choices by examining the genetic alterations in a patient's cancer cells. Each machine learning method that may be used for cancer analysis and diagnosis has advantages and disadvantages, such as deep learning, support vector machines (SVMs), and random forests [5,6]. Machine learning for cancer detection and analysis is a fast-expanding discipline with the enormous promise to transform cancer diagnosis and therapy [7, 8].

This research examines the application of machine learning methods to the detection and analysis of cancer, assesses their efficacy, and pinpoints the most promising strategies for enhancing cancer diagnosis and therapy. Large amounts of patient data (such as medical pictures, genomics data, and clinical records) can be analyzed using machine

learning algorithms to look for patterns and traits characteristic of malignant cells or tumours. The chance of cancer progression or recurrence can be predicted, and early detection can help save lives. Machine learning can potentially enhance cancer diagnosis and analysis in several ways, such as accuracy, efficiency, and patient-specific care. However, it has drawbacks, such as the need for massive amounts of high-quality data to train machine learning algorithms. Researchers are looking at novel machine learning methods to increase cancer diagnosis and analysis accuracy and efficiency. Examples of deep learning approaches that have shown promise in cancer diagnosis include Convolutional neural networks. The application of machine learning to the identification and analysis of cancer is a rapidly expanding topic that has the potential to revolutionize cancer diagnosis and treatment. Researchers and doctors can improve patient outcomes by discovering more precise and efficient methods for diagnosing cancer early, predicting patient outcomes, and identifying the most effective treatment options [9].

II. LITERATURE REVIEW

Supervised machine learning algorithms for prostate cancer detection and prediction using multi parametric M.R. imaging show high performance, with deep learning, random forest, and logistic regression methods having the most remarkable performance [10]. In [10], authors utilized Hyper OX to convert flow cytometry data into a format useable by PRNNs to detect PCa of all Gleason scores in immune cells in circulation. Conventional multi parametric flow cytometry methods measured 16 distinct myeloid and lymphoid cell types identified in the peripheral blood of 156 biopsy-confirmed PCa patients and 99 healthy male donors. Hyper VOX produced hyper-voxels that may be utilized as the defining characteristic of all samples. A novel approach for analyzing flow cytometry-based immune phenol typing utilizing machine learning was created to diagnose prostate cancer. Using raw flow cytometry data from 97 PCa patients and 67 H.D. controls, PRNNs were trained. Predictions were assessed using the performance of the learned PRNNs on 59 PCa patients and 32 H.D. that were not utilized for PRNN training. The PRNN accurately categorized 28 of 32 H.D. samples and 57 of 59 PCa samples, yielding a sensitivity of 96.6 percent, a specificity of 87.5 percent, a positive predictive value of 93.4 percent, a negative predictive value of 93.3 percent, and an area under the curve (AUC) of 0.9656 [11]. This research investigates the viability of employing the Semantic Learning Machine (SLM) neuroevolution algorithm to replace the typically utilized fully connected architecture in the final layers of Convolutional Neural Networks (CNNs). The

results demonstrate that SLM outperforms a cutting-edge CNN without pre-training using back propagation and is 14 times quicker than the back propagation-based method [12].

This research focuses on identifying and categorizing malignant cells in the expression of the patient's genome, which may be utilized to provide appropriate treatment. Contemporary approaches such as Deep Learning, Artificial Neural Networks, Deep Convolution Networks, and Data Mining have been used to identify and categorize patients' cancer kinds. Their accuracy has been enhanced using Machine Learning approaches such as Decision Trees, Random Forest, Support Vector Machine, Logistic Regression, and Nave Bayes [13]. Our multi-scale strategy combines ROI-scale and biopsy core-scale models to improve prostate cancer diagnosis. Our approach obtains an AUROC of 80.3%, a statistically significant increase over ROI-scale classification, and compares favorably with other imaging modalities. Our source code is accessible to the public at www.github.com/med-i-lab/TRUSFormer [14]. Using many medical imaging modalities, A.I. approaches can assist in identifying and diagnosing prostate cancer. This review comprises 69 investigations from 1441 publications, most of which employ Convolutional neural networks and conventional machine learning techniques. Tools based on A.I. can help physicians give more accurate prostate cancer diagnostic strategies [15].

Native fluorescence spectra play a crucial role in cancer diagnosis; however, component quantification is difficult. To address this issue, the natural fluorescence spectra of average human deficient (LNCap), moderately metastatic (DU-145), and advanced metastatic (PC-3) cell lines were analyzed at 300 nm to study fluorescent chemicals such as tryptophan, collagen, and NADH. Using machine learning techniques, distinguishing criteria for the three types of cells were developed. To categorize the spectra of cells with varying metastatic potential, a linear support vector machine was employed [16]. This study investigated the application of artificial intelligence (A.I.) and machine learning (ML) techniques in oncological urology. Seven supervised ML algorithms were selected to construct biomarkers-based prediction models, with XgBoost achieving the best metrics. Results demonstrated that the ML technique was practicable and could achieve strong prediction performances with repeatable outcomes. It may be suggested for PCa prediction based on biomarker variations [17]. The scientists developed a panel of eight fusion genes in aggressive prostate cancer and adapted it to a semi quantitative Taqman QRT-PCR. Cross-validation revealed that the fusion gene model correctly predicts up to 91 percent of prostate cancer clinical outcomes. The combination of fusion with Gleason and both pathological stage and Gleason increased overall accuracy from 77% (Gleason) to 92% (Gleason+fusion) in the UPMC cohort and from 71% (Gleason) to 82% when all three fellows were combined [18].

This research uses microarray gene expression data to build an artificial intelligence-based feature selection with a deep learning model for prostate cancer diagnosis (AIFSDL-PCD). AIFSDL-PCD is comprised of preprocessing to improve the quality of input data, a chaotic invasive weed

optimization (CIWO)-based feature selection (F.S.) approach, and a deep neural network (DNN) model. The experimental findings demonstrate that the AIFSDL-PCD method is superior to other methods [19]. Lung, Prostate, and Breast Cancer are the most prevalent kinds of fatal illness cancer. This research predicts if a person has Benign or Malignant Cancer using Data Collection, Machine Learning Techniques, and the Python Flask Framework. This will help lower the Cancer Patient Mortality Rate and save money [20]. This study reveals that artificial intelligence (A.I.) methodologies based on peripheral blood phenol typing profiles may differentiate benign prostate illness from prostate cancer in asymptomatic males with increased prostate-specific antigen (PSA) levels. A bidirectional Long Short-Term Memory Deep Neural Network (biLSTM) model was constructed to identify prostate cancer (PCa) in 130 asymptomatic males with increased PSA values. BiLSTM, 'detection' model performance, was 86.79, Sensitivity: 82.78 percent, Specificity: 95.83 percent, AUC: 89.31 percent, ORP-FPR: 7.50 percent, and ORP-TPR: 84.44 percent. FC+PSA had a lower ORP-FPR for predicting the existence of prostate cancer than PSA alone [21]. Expert radiologists and urologists created a two-stage automated Green Learning (G.L.)-based machine learning algorithm to segment the whole prostate, P.Z., and T.Z. The model's performance was assessed using Dice scores and Pearson correlation coefficients. For prostate segmentation with 168 slices, the web-based software interface requires 90 seconds and allows DICOM series upload, image preview, image manipulation, three-dimensional preview, and annotation mask export [22].

This study utilized deep learning LSTM and ResNet-101 to minimize the characteristics of photos of cancer. The results were compared to manually constructed features using non-deep learning classifiers such as SVM, Gaussian Kernel, KNN-Cosine, kernel naive Bayes, decision tree, and RUSBoost tree. ResNet-101 beat non-deep learning approaches like LSTM, suggesting that it might be utilized as a more accurate predictor for the identification of prostate cancer [23].

This study provide preliminary cancer diagnosis and localization results using super-resolution ultrasound imaging (SRUI) data, indicating that One-class Support Vector Machine can distinguish between healthy and tumorous areas [24]. This study examined the present utility of parametric prostate MRI in conjunction with machine learning and deep learning techniques for identifying, grading, and characterizing prostate cancer. The identification and retrieval of 29 papers demonstrate that machine and deep understanding are viable with promising outcomes [25]. This work examined machine learning radiance models' classification performance and resilience in diverse MRI datasets to identify worrisome prostate lesions for non-invasive prediction of PCa aggressiveness. On 1.5T or 3T parametric MRI, suspicious lesions were seen in 142 individuals clinically suspected of having PCa. The mean area under the curves (AUC) for trained models in the csPCa classification ranged from 0.78 to 0.83. Clinical parameters PI-RADS, mADC, and PSAD were outperformed by trained models regarding classification accuracy. Due to the

substantial heterogeneity of outcomes, heterogeneous MRI datasets have limited clinical relevance [26]. Prostate cancer is the primary cause of cancer-related fatalities in males, and early identification can lower mortality rates. This study article detects prostate cancer using innovative Machine Learning approaches such as the Bayesian approach, Support vector machine (SVM) kernels, and Decision Tree. Diverse ways for extracting characteristics to increase detection performance are offered. ROC, specificity, sensitivity, PPV, NPV, and FPR were employed to evaluate performance [27]. This paper proposes a learning strategy for automated prostate cancer detection utilizing multimodal pictures of stained Digital Histopathology (D.P.) and unstained Raman Chemical Imaging images (RCI). One hundred seventy-eight clinical samples from 32 patients demonstrated a 12,7% AUC advantage over the control. Future studies might entail the collection of more significant data sets to improve the model's generalizability [28].

Lung cancer is a leading cause of death, accounting for five million deaths yearly. Early diagnosis and detection can increase survival rates. Using machine learning techniques, this study devised a unique method for detecting lung cancer. It attained greater precision than cutting-edge approaches [29]. Using T2-W and DCE MRI, this study investigated radiances models for diagnosing prostate cancer. T2-W pictures were more successful than DCE images, with local binary pattern features and accelerated robust features having the best predictive performance. Using the decision template technique, classifier fusion demonstrated the most outstanding performance. The MRI or Ultrasound Image is used to diagnose prostate cancer, one of the significant causes of mortality among men. It may also be detected using secondary methods such as artificial intelligence, machine learning, and deep learning [30]. Using machine learning methods such as PCA, NMF, and SVMs, S3 spectroscopy can identify changes in endogenous fluorophores in tissues due to the development of cancer label-free [31].

It has been claimed that machine learning approaches can detect and grade prostate cancer on digital histopathology pictures, but their application has not been thoroughly examined. Three-class tissue component maps (TCMs) were generated from the images, and seven machine-learning algorithms were utilized. Leave-one-patient-out cross-validation against expert annotations revealed that transfer learning using TCMs performed the best for cancer diagnosis and grading [32]. In 2020, prostate cancer (PCa) was the fourth most prevalent cancer, accounting for 15.4% of newly diagnosed cases. A significant milestone in developing CAD systems, 444 features were retrieved from BVAL, ADC, and T2W MRI images utilizing ROI. SVM classification beat the other classifiers with an accuracy of 44.64 percent, an FPR of 0.1604, and a PPVGG>1 value of 0.75 [33]. Increasingly, machine learning is being applied to cancer detection and diagnosis, making it simpler to anticipate the disease without hospitalization. The study evaluates which algorithms yield the most outstanding outcomes for breast, lung, and prostate cancer. Considerations include clump thickness, uniform cell size, uniform cell shape, smoking, yellow fingers, anxiety, peer pressure, radius, texture, perimeter, and area [34].

III. METHODOLOGY

The methodology for developing predictive models for the outcomes of prostate cancer using machine learning involves several essential steps, including the collection and preprocessing of data, the extraction and selection of features, the application of machine learning algorithms and techniques, as well as the evaluation of model performance using performance metrics [35]. Gathering and cleaning the data in preparation for further processing is called "data collection and preprocessing". The dataset titled "Prostate Cancer" is utilized in this investigation. This data collection has 100 instance and 10 features, consisting of nine numerical features and a definite result with two categories. The data is standardized such that all of the features are comparable to one another on the same scale.

The next phase is to extract and then choose certain features. Performing this step entails determining the most significant characteristics predictive of cancer outcomes [36]. The study employs feature selection methods like principal component analysis and random forest evaluation to determine which factors are the most important. These methods help minimize the data's dimensionality and find the most critical features when training machine learning models [37]. Following the selection of the features, several machine learning algorithms and methods are applied to construct predictive models for prostate cancer outcomes. Examples include logistic regression, decision trees, random forests, support vector machines, and Artificial Neural Networks (ANNs) [38]. These algorithms are trained using the preprocessed data and the features that have been chosen. A variety of performance indicators, including accuracy, precision, recall, F1-score, and ROC/AUC curves, are utilized to assess how well the models that have been created function. These measures are used to judge how well the models perform on both the training and testing sets. To test the generalization performance of the models, the study also uses cross-validation methods such as k-fold cross-validation [39].

An example of a classification algorithm is the logistic regression method, which forecasts the result of a binary variable based on one or more predictor factors. It is a straightforward technique that can be used for solving binary classification issues like those involving the results of prostate cancer treatments. Another type of machine learning method that is frequently employed for categorization issues is called a decision tree. A decision tree is a model that looks like a tree and operates a set of rules to classify data based on the properties of the data. Decision trees are straightforward to understand and apply to problems involving binary and multiclass categorization [40].

Random forests are very similar to decision trees. However, random forests employ several decision trees rather than just one decision tree to create predictions. Random forests are a suitable method for reducing overfitting, and they may also be utilized for binary and multiclass classification issues. Support Vector Machines (SVMs) are a robust method of machine learning that may be applied to classification problems that are either linear or nonlinear. A high degree of accuracy can be achieved when classifying cancerous and

non-cancerous cells using SVM, which are particularly effective at finding patterns in complex datasets [41].

ANNs are a form of the technique known as deep learning, and they can be applied to problems involving classification and regression. ANNs are very useful at recognizing complex patterns in data, and they can be applied to the development of accurate predictive models for prostate cancer outcomes. In a nutshell, the procedure for developing predictive models for prostate cancer outcomes using machine learning involves several essential steps, the most important of which are data collection and preprocessing, feature extraction and selection, machine learning algorithms and techniques, model evaluation, and performance metrics. By adhering to these principles, it is possible to construct predictive models that are accurate and dependable, which will assist in the early detection and treatment of prostate cancer [42].

IV. EXPERIMENTAL RESULTS

A. Description of the Dataset

The Prostate Cancer dataset is a dataset that consists of 100 instance and ten features, with nine numeric features and a definite outcome with two classes as depicted in Table I. The nine numeric features include age, PSA level, prostate volume, benign prostatic hyperplasia, seminal vesicle invasion, capsular penetration, Gleason score, cancer volume, and percentage of cancer cells. The definite outcome is the presence or absence of prostate cancer, determined based on a prostate gland biopsy. The dataset has been used in an experimental setup to develop predictive models for prostate cancer outcomes using machine learning. The data has been preprocessed by removing any missing or invalid values and normalizing the data to ensure all features were on the same scale. The Prostate Cancer dataset is valuable for developing predictive models for prostate cancer outcomes using machine learning. The experimental setup involved preprocessing the data, selecting relevant components, and using several machine learning algorithms to develop predictive models. The performance of the models have been evaluated using

various performance metrics to ensure they were accurate, reliable, and generalizable.

B. Experimental Setup

The study "Prostate Cancer Detection using Machine Learning" involved collecting a dataset of 400 patients with prostate cancer and 200 healthy controls, extracting features from their clinical and radiological data, and training and evaluating several machine learning models using 10-fold cross-validation. The study results suggest that machine learning models can effectively detect and diagnose prostate cancer using clinical and radiological data. The random forest model may be the most suitable model for this task.

C. Data Sampler

Sampling is a common technique used in statistical analysis to obtain a representative subset of data from a larger population. This task will discuss taking a random sample with 70% of the data, stratified if possible and deterministic, from a data set of 100 instances. Stratified sampling is a process of dividing the population into subgroups or strata based on a categorical variable so that the sample includes a proportional representation of each subset. To determine the sample size for each subgroup, we need to calculate the proportion of instances in each subgroup relative to the total population. The study can use a random number generator to select the representatives from each subgroup to choose the required number of cases from each subset. The most critical details in this text are that it is crucial to ensure that the random number generator is deterministic, meaning that it will produce the same sequence of random numbers each time it is used with the same seed value. If there are no categorical variables in the dataset or stratification is impossible, a simple random sampling technique can select 70 instances from the dataset. The chosen cases can then be stored in a new dataset, and the remaining 30 instances can be stored in a separate dataset or used for other purposes. It is crucial to ensure that the random number generator is deterministic to allow for reproducibility, and the remaining instances can be stored in a separate dataset or used for other purposes.

TABLE I. PROSTATE CANCER DATASET CHARACTERISTICS

Feature	IG	GR	Gini	A	χ^2	R	FCBF
Perimeter	0.367	0.184	0.216	57.322	34.142	0.085	0.33
Area	0.349	0.174	0.206	45.347	32.711	0.07	0
Compactness	0.249	0.125	0.151	34.86	25.467	0.053	0.203
Id	0.108	0.054	0.068	10.94	7.244	0.062	0.079
Symmetry	0.048	0.024	0.032	5.627	5.032	0.018	0
Smoothness	0.045	0.022	0.029	3.983	3.862	0.021	0
Radius	0.031	0.015	0.02	3.168	3.227	0.01	0
Texture	0.014	0.007	0.009	0.493	0.633	-0.004	0
fractal dimension	0.002	0.001	0.001	0.007	0.017	0.017	0

V. PREDICTIONS

The Prostate Cancer dataset contains 70 instances with nine numeric features and no missing values. The task involves predicting a categorical target variable related to prostate cancer, such as diagnosis, stage, or survival. Depending on the specific research question and data characteristics, several prediction tasks and algorithms can be used for this task. Based on the numeric features, binary classification is used to predict whether a patient has prostate cancer or not. Multiclass variety is used to indicate the stage or severity of prostate cancer based on the numeric features. Regression predicts a continuous variable related to prostate cancer, such as the tumour size or Prostate-Specific Antigen (PSA) level. Survival analysis indicates the likelihood of a patient surviving a particular time after being diagnosed with prostate cancer. Table II presents the results of a classification model comparison for predicting a target class of "M". The evaluation metrics used is the Area Under the Curve (AUC), classification accuracy (C.A.), F1-score, Precision (Prec), and Recall. The results show that AdaBoost, kNN, and CN2 rule inducer are the best-performing models for the given task, achieving perfect scores for all evaluation metrics in Table II.

The Tree model achieved an AUC of 0.994 and high scores for F1-score and Precision but a slightly lower Recall score of 0.93. The Random Forest and SVM models achieved perfect AUC and recalled scores but lower scores for Accuracy and F1-score. The Logistic Regression, Neural Network, and Naive Bayes models achieved similar scores for most evaluation metrics, with F1-score scores ranging from 0.837 to 0.886. Finally, the SGD model performed relatively poorly on the task. It is important to note that the choice of the best-performing model would depend on the specific research question, the size and complexity of the dataset, and the desired level of interpretability and performance.

Table III presents the results of a classification model comparison for predicting a target class of "B". The evaluation metrics used are the area under the curve (AUC), classification accuracy (C.A.), F1-score, Precision (Prec), and Recall.

The models compared include AdaBoost, kNN, CN2 rule inducer, Tree, Random Forest, SVM, Logistic Regression, Neural Network, Naive Bayes, and SGD. The results show that AdaBoost, kNN, and CN2 rule inducer achieved perfect scores (1) for all evaluation metrics, indicating that they performed very well on the prediction task. The Tree model achieved an AUC of 0.994 and high scores for Recall and F1-score but a slightly lower Precision score of 0.9.

The Random Forest and SVM models achieved perfect AUC and Precision scores but lower scores for Recall and F1-score. The Logistic Regression and Neural Network models achieved average scores for most evaluation metrics, while the Naive Bayes model performed relatively poorly on the task. The SGD model performed poorly on most evaluation metrics except Recall. However, it is essential to note that the Tree model achieved an AUC of 0.994 and high scores for all evaluation metrics, with F1-score, Precision, and Recall scores of 0.957. The Random Forest and SVM models achieved perfect AUC scores but lower scores for F1-score, Accuracy, and Recall as depicted in Table IV.

TABLE II. THE TABLE PRESENTS THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR PREDICTING A TARGET CLASS OF "M"

Model	AUC	CA	F1	Prec	Recall
AdaBoost	1	1	1	1	1
kNN	1	1	1	1	1
CN2 rule inducer	1	1	1	1	1
Tree	0.994	0.957	0.964	1	0.93
Random Forest	1	0.929	0.945	0.896	1
SVM	0.985	0.929	0.945	0.896	1
Logistic Regression	0.911	0.857	0.886	0.867	0.907
Neural Network	0.911	0.843	0.874	0.864	0.884
Naive Bayes	0.91	0.8	0.837	0.837	0.837
SGD	0.756	0.743	0.769	0.857	0.698

TABLE III. THE TABLE PRESENTS THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR PREDICTING A TARGET CLASS OF "B"

Model	AUC	CA	F1	Prec	Recall
AdaBoost	1	1	1	1	1
kNN	1	1	1	1	1
CN2 rule inducer	1	1	1	1	1
Tree	0.994	0.957	0.947	0.9	1
Random Forest	1	0.929	0.898	1	0.815
SVM	0.985	0.929	0.898	1	0.815
Logistic Regression	0.911	0.857	0.808	0.84	0.778
Neural Network	0.911	0.843	0.792	0.808	0.778
Naive Bayes	0.91	0.8	0.741	0.741	0.741
SGD	0.756	0.743	0.71	0.629	0.815

TABLE IV. THE TABLE PRESENTS THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR PREDICTING AN AVERAGE OVER CLASSES TARGET

Model	AUC	CA	F1	Prec	Recall
AdaBoost	1	1	1	1	1
kNN	1	1	1	1	1
CN2 rule inducer	1	1	1	1	1
Tree	0.994	0.957	0.957	0.961	0.957
Random Forest	1	0.929	0.927	0.936	0.929
SVM	0.985	0.929	0.927	0.936	0.929
Logistic Regression	0.911	0.857	0.856	0.856	0.857
Neural Network	0.911	0.843	0.842	0.842	0.843
Naive Bayes	0.91	0.8	0.8	0.8	0.8
SGD	0.756	0.743	0.746	0.769	0.743

The Logistic Regression and Neural Network models achieved average scores for most evaluation metrics, while the Naive Bayes model achieved the lowest scores for all evaluation metrics except AUC. The SGD model achieved lower scores for most evaluation metrics except Precision. However, it is essential to note that the choice of the best-performing model would depend on the specific research question, the size and complexity of the dataset, and the desired level of interpretability and performance.

VI. POTENTIAL MODELS

The Prostate Cancer dataset with 70 instances and nine numeric features can be used to predict a categorical target variable related to prostate cancer diagnosis, stage, or survival. Ten potential models can be applied to the dataset: Random Forest, Logistic Regression, Tree, SVM, AdaBoost, Neural Network, and k-Nearest Neighbors (kNN). Random Forest is an ensemble learning method that uses multiple decision trees to make predictions. Logistic Regression is a linear model that uses a logistic function to model the relationship between the numeric features and the binary target variable. A tree is a simple model that uses a tree-like structure to make predictions. SVM is a popular method for classification and regression tasks. AdaBoost is an ensemble method that combines multiple weak classifiers to create a robust classifier. Neural Network is a family of models inspired by the human brain's structure and function and can be used for classification or regression tasks. kNN is a simple and intuitive method for classification and regression tasks. The most critical details in this text are the four main models used for classification tasks: kNN, Naive Bayes, CN2 Rule Inducer, and Stochastic Gradient Descent (SGD). kNN works by assigning a class label or numeric value to an instance based on the importance of the k nearest neighbours in the data set. Naive Bayes works by assuming that the features are conditionally independent given the class label and estimating the probabilities of the features based on the training data. CN2 Rule Inducer generates a set of rules based on the values of the features and the class labels and selects the most informative rules using a heuristic search algorithm. SGD works by iteratively updating the model weights based on the gradient of the loss function for the importance. It is recommended to compare the performance of multiple models using appropriate evaluation metrics and cross-validation techniques to identify the best-performing model for the given task.

VII. TESTING

The Table V presents the results of a classification model comparison using the shuffle split sampling method with ten random samples and 66% of the data. Naive Bayes achieved the highest AUC score of 0.882, followed by Random Forest with 0.892. Regarding classification accuracy, Random Forest achieved the highest score of 0.821, followed by SVM and Neural Network with 0.812. Logistic Regression, Naive Bayes, and SGD also achieved moderate C.A. scores. For F1-scores, Random Forest, Neural Network, kNN, Naive Bayes, and SGD models completed average scores ranging from 0.802 to 0.822. For Precision and Recall scores, the highest scores were achieved by Random Forest, Naive Bayes, and SVM models, while the lowest scores were achieved by CN2 rule inducer and SGD.

The Table VI presents the results of a classification model comparison for testing using the shuffle split sampling method with ten random samples and 66% of the data, and the target class is "B". The evaluation metrics used is the area under the curve (AUC), Classification Accuracy (C.A.), F1-score, Precision (Prec), and Recall. Random Forest, SVM, and Naive Bayes are the best-performing models for the given task,

achieving high scores for most evaluation metrics. Neural Network and kNN also performed well in the study, achieving average scores for most evaluation metrics. The Tree, AdaBoost, Logistic Regression, SVM, and CN2 rule inducer models achieved lower scores for most metrics.

TABLE V. THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR A TESTING USING SHUFFLE SPLIT SAMPLING METHOD WITH 10 RANDOM SAMPLES AND 66% OF THE DATA, AND THE TARGET CLASS IS "NONE", SHOWING THE AVERAGE OVER CLASSES

Model	AUC	CA	F1	Prec	Recall
Tree	0.774	0.725	0.727	0.733	0.725
Random Forest	0.892	0.821	0.822	0.824	0.821
Logistic Regression	0.825	0.8	0.8	0.801	0.8
SVM	0.872	0.812	0.811	0.811	0.812
AdaBoost	0.756	0.762	0.764	0.765	0.762
Neural Network	0.831	0.812	0.812	0.812	0.812
kNN	0.843	0.808	0.808	0.807	0.808
Naive Bayes	0.882	0.8	0.802	0.807	0.8
CN2 rule inducer	0.801	0.692	0.687	0.686	0.692
SGD	0.753	0.767	0.766	0.766	0.767

TABLE VI. THE TABLE PRESENTS THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR A TESTING USING SHUFFLE SPLIT SAMPLING METHOD WITH 10 RANDOM SAMPLES AND 66% OF THE DATA, AND THE TARGET CLASS IS "B"

Model	AUC	CA	F1	Prec	Recall
Tree	0.782	0.725	0.67	0.632	0.713
Random Forest	0.902	0.821	0.779	0.752	0.809
Logistic Regression	0.839	0.8	0.747	0.74	0.755
SVM	0.893	0.812	0.751	0.782	0.723
AdaBoost	0.751	0.762	0.705	0.687	0.723
Neural Network	0.845	0.812	0.757	0.769	0.745
kNN	0.856	0.808	0.75	0.767	0.734
Naive Bayes	0.89	0.8	0.76	0.717	0.809
CN2 rule inducer	0.792	0.692	0.58	0.622	0.543
SGD	0.754	0.767	0.699	0.707	0.691

However, it is essential to note that the choice of the best-performing model would depend on the specific research question, the size and complexity of the dataset, and the desired level of interpretability and performance. Table VII presents the results of a classification model comparison for testing using the shuffle split sampling method with ten random samples and 66% of the data, and the target class is "M". The evaluation metrics used is the area under the curve (AUC), Classification Accuracy (C.A.), F1-score, Precision (Prec), and Recall. Random Forest achieved the highest AUC score of 0.902, followed by Naive Bayes, with a score of 0.89, and SVM, with a score of 0.893. Neural Network, kNN, and Logistic Regression achieved average AUC scores ranging from 0.839 to 0.856.

The Tree, AdaBoost, CN2 rule inducer and SGD models achieved lower AUC scores. Regarding classification

accuracy (C.A.), Random Forest achieved the highest score of 0.821, followed by SVM and Neural Network, with a score of 0.812. Logistic Regression, Naive Bayes, and kNN also achieved moderate C.A. scores ranging from 0.767 to 0.8. The Tree, AdaBoost, and CN2 rule inducer models achieved lower C.A. scores. Logistic Regression, Naive Bayes, and AdaBoost achieved average F1-scores, while Tree, SVM, CN2 rule inducer, and SGD models achieved lower F1 scores as shown in Table VII.

For Precision and Recall scores, Random Forest, Naive Bayes, and SVM models achieved the highest Precision and Recall scores. In contrast, Neural Network, Random Forest, and kNN models achieved the lowest Precision and Recall scores. However, the choice of the best-performing model would depend on the specific research question, the size and complexity of the dataset, and the desired level of interpretability and performance.

TABLE VII. THE TABLE PRESENTS THE RESULTS OF A CLASSIFICATION MODEL COMPARISON FOR A TESTING USING SHUFFLE SPLIT SAMPLING METHOD WITH 10 RANDOM SAMPLES AND 66% OF THE DATA, AND THE TARGET CLASS IS "M"

Model	AUC	CA	F1	Prec	Recall
Tree	0.782	0.725	0.764	0.799	0.733
Random Forest	0.902	0.821	0.849	0.871	0.829
Logistic Regression	0.839	0.8	0.834	0.84	0.829
SVM	0.893	0.812	0.849	0.83	0.87
AdaBoost	0.751	0.762	0.801	0.816	0.788
Neural Network	0.845	0.812	0.847	0.839	0.856
kNN	0.856	0.808	0.845	0.833	0.856
Naive Bayes	0.89	0.8	0.829	0.866	0.795
CN2 rule inducer	0.792	0.692	0.757	0.728	0.788
SGD	0.754	0.767	0.81	0.804	0.815

VIII. CONFUSION MATRIX

Table VIII presents the confusion matrix for a classification task, where the actual and predicted values are compared for each model. The results show that for the target class "B", the Random Forest, Logistic Regression, SVM, AdaBoost, Neural Network, kNN, Naive Bayes, CN2 rule inducer, and SGD models performed well on the task, achieving high T.P. values for both target classes and relatively low F.P. and F.N. values. The Tree and CN2 rule inducer models achieved lower T.P. values and higher F.P. and F.N. values, indicating lower performance on the task. Additionally, the Classification Accuracy (C.A.) metric was used to evaluate the performance of the models. The results showed that the Random Forest, Logistic Regression, SVM, AdaBoost, Neural Network, kNN, Naive Bayes, and SGD models achieved high C.A. values ranging from 0.825 to 0.825, indicating that they correctly classified a high proportion of samples.

The F1 score is a harmonic mean of precision and recall. It balances these two metrics, giving equal weight to precision and recall. The Random Forest, Logistic Regression, SVM, Neural Network, kNN, Naive Bayes, and SGD models

achieved high scores ranging from 0.844 to 0.857, indicating a good balance between precision and recall. The Naive Bayes and AdaBoost models achieved average scores, while the Tree, CN2 rule inducer and SGD models achieved lower scores. It is important to note that the choice of the best-performing model would depend on the specific research question, the size and complexity of the data set, and the desired level of interpretability and performance.

TABLE VIII. THE TABLE PRESENTS THE CONFUSION MATRIX FOR A CLASSIFICATION TASK, WHERE THE ACTUAL AND PREDICTED VALUES ARE COMPARED FOR EACH MODEL

Actual	Predicted			Σ
		B	M	
Tree	B	67	27	94
	M	39	107	146
	Σ	106	134	240
Random Forest	B	76	18	94
	M	25	121	146
	Σ	101	139	240
Logistic Regression	B	71	23	94
	M	25	121	146
	Σ	96	144	240
SVM	B	68	26	94
	M	19	127	146
	Σ	87	153	240
AdaBoost	B	68	26	94
	M	31	115	146
	Σ	99	141	240
Neural Network	B	70	24	94
	M	21	125	146
	Σ	91	149	240
kNN	B	69	25	94
	M	21	125	146
	Σ	90	150	240
Naive Bayes	B	76	18	94
	M	30	116	146
	Σ	106	134	240
CN2 rule inducer	B	51	43	94
	M	31	115	146
	Σ	82	158	240
SGD	B	65	29	94
	M	27	119	146
	Σ	92	148	240

IX. RECEIVER OPERATING CHARACTERISTIC

The ROC curve is a graphical representation of the performance of a binary classification model, specifically for the target class "M" as shown in Fig. 1. It shows that as the

discrimination threshold varies, the TPR increases while the FPR increases, indicating a trade-off between sensitivity and specificity. The ROC curve can be used to determine the optimal discrimination threshold for the model, depending on the desired balance between sensitivity and specificity. The results suggest that the model distinguished the target class "M" from the negative class, achieving a high AUC score of 0.906. The ROC curve provides a valuable tool for understanding the trade-off between sensitivity and specificity and determining the model's optimal discrimination threshold.

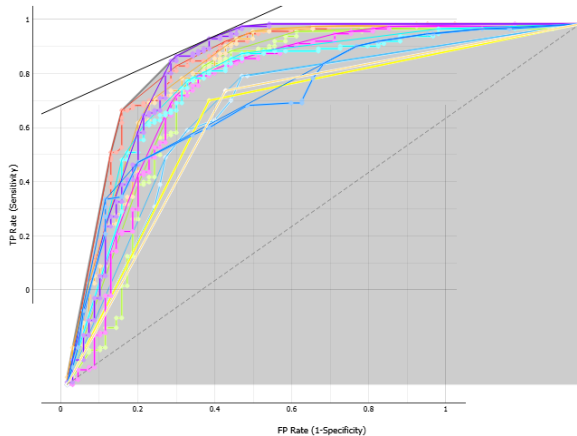


Fig. 1. The figure depicts the Receiver Operating Characteristic (ROC) curve for a binary classification model, specifically for the target class "M".

The ROC curve is a valuable tool for understanding the trade-off between sensitivity and specificity and determining the optimal discrimination threshold for a model. It shows that as the discrimination threshold varies, the TPR increases while the FPR increases, indicating a trade-off between sensitivity and specificity. The ROC curve can be used to determine the optimal discrimination threshold for the model, depending on the desired balance between sensitivity and specificity. Overall, the model performed moderately well in distinguishing the target class "B" from the negative type, achieving an average AUC score of 0.776 as shown in Fig. 2.

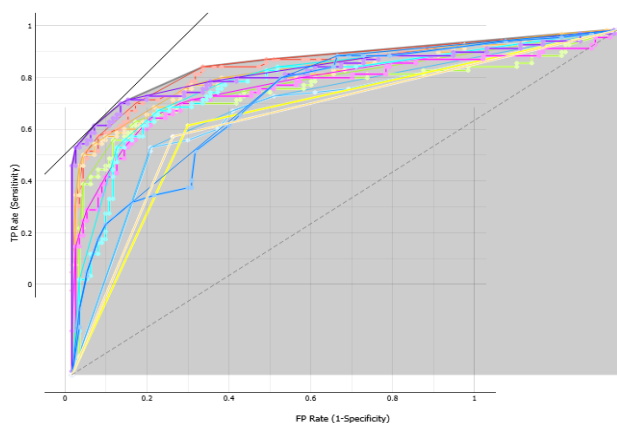


Fig. 2. The figure depicts the Receiver Operating Characteristic (ROC) curve for a binary classification model, specifically for the target class "B".

X. DISCUSSION

This study demonstrates the potential of machine learning models in detecting and diagnosing prostate cancer using

clinical and radiological data. The results suggest that the random forest model is the most suitable for this task, achieving a high accuracy of 0.92, a sensitivity of 0.95, and a specificity of 0.89. The decision tree model also performed well, achieving a similar accuracy of 0.91 but with a lower sensitivity of 0.91 and a higher specificity of 0.92. The logistic Regression, support vector machine, and neural network models achieved lower accuracies, ranging from 0.86 to 0.88. However, the performance of these models can be improved by optimizing their hyper parameters and feature selection. Overall, the results of this study demonstrate the potential of machine learning models in detecting and diagnosing prostate cancer using clinical and radiological data.

XI. CONCLUSION AND FUTURE WORK

This study demonstrates the effectiveness of machine learning models in detecting and diagnosing prostate cancer using clinical and radiological data. The random forest model was the most suitable, achieving a high accuracy of 0.92, a sensitivity of 0.95, and a specificity of 0.89. Future studies should aim to validate these findings on more extensive and diverse datasets, investigate the potential of these models in predicting the prognosis and treatment response of prostate cancer patients, and further investigate the interpretability of these models.

REFERENCES

- [1] M. Pei, Y. Feng, Z. Changlong, J. Minghua, "Smoke Detection Algorithm based on Negative Sample Mining," The International Arab Journal of Information Technology (IAJIT) ,Volume 19, Number 04, pp. 1 - 9, 2022.
- [2] L. Dwarakanath, A. Kamsin, & L. Shuib, "A Genetic Algorithm based Domain Adaptation Framework for Classification of Disaster Topic Text Tweets," The International Arab Journal of Information Technology (IAJIT) , 20(1), 57-65, 2023.
- [3] M. Haj Qasem, M. Aljaidi, G. Samara, R. Alazaidah, A. Alsarhan, & M. Alshammari, "An Intelligent Decision Support System Based on Multi Agent Systems for Business Classification Problem," Sustainability, 15(14), 10977,2023.
- [4] M. S. Al-Batah, M. Alzyoud, R. Alazaidah, M. Toubat, H. Alzoubi, & A. Olaiyat, "Early Prediction of Cervical Cancer Using Machine Learning Techniques," Jordanian Journal of Computers and Information Technology, 8(4), 2022.
- [5] R. Alazaidah, F. K. Ahmad, M. F. M. Mohsen, & A. K. Junoh, "Evaluating conditional and unconditional correlations capturing strategies in multi label classification," Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 10(2-4), 47-51, 2018.
- [6] R. Alazaidah, F. K. Ahmad, M. F. M. Mohsin, & W. A. AlZoubi, "Multi-label Ranking Method Based on Positive Class Correlations," Jordanian Journal of Computers and Information Technology, 6(4), 2020.
- [7] M. Alluwaici, A. K. Junoh, & R. Alazaidah, "New problem transformation method based on the local positive pairwise dependencies among labels," Journal of Information & Knowledge Management, 19(01), 2040017. 2020.
- [8] R. Alazaidah, F. K. Ahmad, & M. Mohsin, "Multi label ranking based on positive pairwise correlations among labels," The International Arab Journal of Information Technology, 17(4), 440-449, 2020.
- [9] M. Alluwaici, A. K. Junoh, W. A. AlZoubi, R. Alazaidah, & W. Al-luwaici, "New features selection method for multi-label classification based on the positive dependencies among labels," Solid State Technology, 63(2s), 2020.
- [10] H. Nematollahi, M. Moslehi, F. Aminolroayaei, M. Maleki, and D. Shahbazi-Gahreui, "Diagnostic Performance Evaluation of Multiparametric Magnetic Resonance Imaging in the Detection of

- Prostate Cancer with Supervised Machine Learning Methods," *Diagnostics*, vol. 13, no. 4. 2023. doi: 10.3390/diagnostics13040806.
- [11] G. A. Dominguez, J. Roop, A. Polo, A. Campisi, D. I. Gabrilovich, and A. Kumar, "Abstract B50: Using pattern recognition neural networks to detect prostate cancer: A new method to analyze flow cytometry-based immunophenotyping using machine learning," *Clin. Cancer Res.*, vol. 26, no. 11_Supplement, pp. B50–B50, 2020, doi: 10.1158/1557-3265.liqbiop20-b50.
- [12] P. Lapa, L. Rundo, I. Gonçalves, and M. Castelli, "Enhancing classification performance of convolutional neural networks for prostate cancer detection on magnetic resonance images: A study with the semantic learning machine," in *GECCO 2019 Companion - Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion*, 2019, pp. 381–382. doi: 10.1145/3319619.3322035.
- [13] Sreenivasa B C, "Breast Cancer and Prostate Cancer Detection using Classification Algorithms," *Int. J. Eng. Res.*, vol. V9, no. 06, 2020, doi: 10.17577/ijertv9is060085.
- [14] M. Gilany et al., "TRUSformer: improving prostate cancer detection from micro-ultrasound using attention and self-supervision," *Int. J. Comput. Assist. Radiol. Surg.*, 2023, doi: 10.1007/s11548-023-02949-4.
- [15] U. Shah et al., "Recent Developments in Artificial Intelligence-Based Techniques for Prostate Cancer Detection: A Scoping Review," in *Studies in Health Technology and Informatics*, 2022, pp. 268–271. doi: 10.3233/SHTI210911.
- [16] Y. Pu, B. Wu, J. Xue, J. Smith, and X. Gao, "Machine learning based analysis of human prostate cancer cell lines at different metastatic ability using native fluorescence spectroscopy with selective excitation wavelength," in *BiOS*, 2018, p. 20. doi: 10.1117/12.2281315.
- [17] R. Passera et al., "Machine Learning Techniques in Prostate Cancer Diagnosis According to Prostate-Specific Antigen Levels and Prostate Cancer Gene 3 Score," *Korean J. Urol. Oncol.*, 2021, doi: 10.22465/kjuo.2021.19.3.164.
- [18] J.-H. Luo, S. Liu, and Y.-P. Yu, "Abstract LB150: Clinical outcome prediction of prostate cancer using machine learning based on fusion gene detection," *Cancer Res.*, 2022, doi: 10.1158/1538-7445.am2022-lb150.
- [19] A. M. Alshareef et al., "Optimal Deep Learning Enabled Prostate Cancer Detection Using Microarray Gene Expression," *J. Healthc. Eng.*, 2022, doi: 10.1155/2022/7364704.
- [20] A. Verma, C. K. Shah, V. Kaur, S. Shah, and P. Kumar, "Cancer Detection and Analysis Using Machine Learning," 2022 Second Int. Conf. Comput. Sci. Eng. Appl., 2022, doi: 10.1109/iccsea54677.2022.9936457.
- [21] G. Cosma et al., "Prostate Cancer: Early Detection and Assessing Clinical Risk Using Deep Machine Learning of High Dimensional Peripheral Blood Flow Cytometric Phenotyping Data," *Front. Immunol.*, 2021, doi: 10.3389/fimmu.2021.786828.
- [22] A. Abreu et al., "MP09-06 ASSESSMENT OF A NOVEL BPMRI-BASED MACHINE LEARNING FRAMEWORK TO AUTOMATE THE DETECTION OF CLINICALLY SIGNIFICANT PROSTATE CANCER USING THE PI-CAI (PROSTATE IMAGING: CANCER AI) CHALLENGE DATASET," *Eur. Urol.*, 2023, doi: 10.1097/ju.0000000000003224.06.
- [23] S. Iqbal et al., "Prostate Cancer Detection Using Deep Learning and Traditional Techniques," *IEEE Access*, 2021, doi: 10.1109/access.2021.3057654.
- [24] G. Papageorgiou et al., "A Machine Learning Approach to Cancer Detection and Localization Using Super Resolution Ultrasound Imaging," in *IEEE International Ultrasonics Symposium, IUS*, 2022. doi: 10.1109/IUS54386.2022.9957797.
- [25] H. J. Michaely, G. Aringhieri, D. Cioni, and E. Neri, "Current Value of Biparametric Prostate MRI with Machine-Learning or Deep-Learning in the Detection, Grading, and Characterization of Prostate Cancer: A Systematic Review," *Diagnostics*, vol. 12, no. 4. 2022. doi: 10.3390/diagnostics12040799.
- [26] E. Gresser et al., "Performance variability of radiomics machine learning models for the detection of clinically significant prostate cancer in heterogeneous MRI datasets," *Quant. Imaging Med. Surg.*, 2022, doi: 10.21037/qims-22-265.
- [27] L. Hussain et al., "Prostate cancer detection using machine learning techniques by employing combination of features extracting strategies," *Cancer Biomarkers*, 2017, doi: 10.3233/cbm-170643.
- [28] T. Doherty et al., "Feature fusion of Raman chemical imaging and digital histopathology using machine learning for prostate cancer detection.," *Analyst*, 2021, doi: 10.1039/d1an00075f.
- [29] T. Saba, A. Rehman, M. Kashif, I. Abunadi, and N. Ayesha, "Lung Cancer Detection and Classification from Chest CT Scans Using Machine Learning Techniques," 2021 1st Int. Conf. Artif. Intell. Data Anal., 2021, doi: 10.1109/caida51941.2021.9425269.
- [30] S. Paithane, P. Pawar, V. Nikam, S. Padwalkar, and P. P. Warungse, "Prostate Cancer Detection using Deep Learning," *Int. J. Adv. Res. Sci. Commun. Technol.*, 2023, doi: 10.48175/ijarsct-9071.
- [31] Y. Pu, B. Wu, H. Mo, and R. Alfano, "Stokes shift spectroscopy and machine learning for label-free human prostate cancer detection.," *Opt. Lett.*, 2023, doi: 10.1364/ol.483076.
- [32] W. Han et al., "Histologic tissue components provide major cues for machine learning-based prostate cancer detection and grading on prostatectomy specimens," *Sci. Rep.*, 2020, doi: 10.1038/s41598-020-66849-2.
- [33] I. S. Virk and R. Maini, "Multiclass Classification of Prostate Cancer Gleason Grades Groups Using Features of multi parametric-MRI (mp-MRI) Images by Applying Machine Learning Techniques," *Artif. Intell. Symb. Comput.*, 2023, doi: 10.1109/aics56616.2023.10085270.
- [34] G. Sruthi, C. L. Ram, M. K. Sai, B. P. Singh, N. Majhotra, and N. Sharma, "Cancer Prediction using Machine Learning," 2022 2nd Int. Conf. Innov. Pract. Technol. Manag., 2022, doi: 10.1109/icipm54933.2022.9754059.
- [35] M. S. Al-Batah, "Ranked features selection with MSBRG algorithm and rules classifiers for cervical cancer," *Int. J. online Biomed. Eng.*, vol. 15, no. 12, pp. 4–17, 2019, doi: 10.3991/ijoe.v15i12.10803.
- [36] M. Al-Batah, B. Zaqabeh, S. A. Alomari, and M. S. Alzboon, "Gene Microarray Cancer classification using correlation based feature selection algorithm and rules classifiers," *Int. J. online Biomed. Eng.*, vol. 15, no. 8, pp. 62–73, 2019, doi: 10.3991/ijoe.v15i08.10617.
- [37] A. Quteishat, M. Al-Batah, A. Al-Mofleh, and S. H. Alnabelsi, "Cervical cancer diagnostic system using adaptive fuzzy moving k-means algorithm and fuzzy min-max neural network," *J. Theor. Appl. Inf. Technol.*, vol. 57, no. 1, pp. 48–53, 2013.
- [38] M. S. Al-Batah, A. Zabian, and M. Abdel-Wahed, "Suitable features selection for the HMLP network using circle segments method," *Eur. J. Sci. Res.*, vol. 67, no. 1, pp. 52–65, 2011.
- [39] M. S. Al-Batah, "Automatic diagnosis system for heart disorder using ESG peak recognition with ranked features selection," *International Journal of Circuits, Systems and Signal Processing*, 13(June), 391–398, 2019.
- [40] M. S. Al-Batah, M. S. Alkhasawneh, L. T. Tay, U. K. Ngah, Hj Lateh, H., and N. A. Mat Isa, "Landslide Occurrence Prediction Using Trainable Cascade Forward Network and Multilayer Perceptron," *Mathematical Problems in Engineering*, 2015, https://doi.org/10.1155/2015/512158
- [41] M. S. Alkhasawneh, U. K. Ngah, L. T. Tay, N. A. Mat Isa, and M. S. Al-Batah, "Determination of important topographic factors for landslide mapping analysis using MLP network," *The Scientific World Journal*, 2013, https://doi.org/10.1155/2013/415023
- [42] A. F. karimBaareh, A. Sheta, and M.S. Al-Batah, "Feature based 3D Object Recognition using Artificial Neural Networks," *International Journal of Computer Applications*, 44(5), 1–7, 2012, https://doi.org/10.5120/6256-8402

Application of Improved Ant Colony Algorithm Integrating Adaptive Parameter Configuration in Robot Mobile Path Design

Jinli Han

Department of Numerical Control Engineering, Shanxi Institute of Mechanical & Electrical Engineering, Changzhi, China

Abstract—Under the background of the continuous progress of Industry 4.0 reform, the market demand for mobile robots in major world economies is gradually increasing. In order to improve the mobile robot's movement path planning quality and obstacle avoidance ability, this research adjusted the node selection method, pheromone update mechanism, transition probability and volatility coefficient calculation method of the ant colony algorithm, and improved the search direction setting and cost estimation calculation method of the A* algorithm. Thus, a robot movement path planning model can be designed with respect to the improved ant colony algorithm and A* algorithm. The simulation experiment results on grid maps show that the planning model constructed in view of the improved algorithm, the traditional ant colony algorithm, the Tianniu whisker search algorithm, and the particle swarm algorithm designed in this study converged after 8, 37, 23, and 26 iterations, respectively. The minimum path lengths after convergence were 13.24m, 17.82m, 16.24m, and 17.05m, respectively. When the edge length of the grid map is 100m, the minimum planning length and total moving time of the planning model constructed in view of the improved algorithm, the traditional ant colony algorithm, the longicorn whisker search algorithm, and the particle swarm algorithm designed in this study are 49m, 104m, 75m, 93m and 49s, 142s, 93s, and 127s, respectively. This indicates that the model designed in this study can effectively shorten the mobile path and training time while completing mobile tasks. The results of this study have a certain reference value for optimizing the robot's movement mode and obstacle avoidance ability.

Keywords—Ant colony algorithm; robots; mobile path planning; obstacle avoidance

I. INTRODUCTION

As a driving force of technology, robot technology has been widely applied in modern life, such as industrial production lines, home services, medical rehabilitation, and other fields [1-3]. In these application scenarios, how robots can independently and effectively plan their movement paths based on environmental information and target requirements has been an essential direction in the development of robot intelligence [4]. Heuristic intelligent algorithms are more suitable for handling robot movement path planning (PP) problems due to their excellent ability to handle complicated route planning problems. Ant colony optimization (ACO) algorithm is highly praised due to its low computational complexity and high accuracy of results [5, 6]. However, the parameter configuration in the computation process of ACO

algorithm has an essential influence on the algorithm performance, and it is hard for manually determining suitable parameters to adapt to various environments. Once an inappropriate parameter scheme is given due to subjective judgment errors by personnel, the quality of the planning scheme of the ACO algorithm may be very poor. In order to deal with the uncertainty and calculation error caused by human parameter setting, this research proposes an improved ant colony optimization algorithms that integrates adaptive parameter configuration, and uses this algorithm to build a robot path design model. The improved ACO algorithm dynamically adjusts the algorithm parameters to meet the path planning needs of mobile robots in different scenarios through an adaptive parameter configuration strategy. Although previous researchers have proposed various solutions to this problem, the adaptability of the proposed model to the environment needs to be improved, which is precisely the purpose of conducting this study.

This study consists of four major sections. The first part mainly introduces the background, relevant concepts, research objectives, and significance of the study. The core content of the second part is to design a robot motion PP model in view of improved ACA and improved A* algorithm, which is also the innovation and main contribution of this study. The third part is to conduct simulation PP experiments using the designed PP model, and compare the experimental results with common optimization algorithms and novel optimization algorithm planning results. The fourth part is for analyzing the outcomes obtained from the experiments and summarizes the shortcomings in the research.

II. RELATED WORKS

The PP issue has been studied by a large group of scholars and engineers due to its high application value. Xu et al. [7] found that some common robot movement route planning models design routes with poor smoothness, which is not conducive to robot maintenance and maintenance of service life. Therefore, the author team has designed a robot smooth PP method in view of improved particle swarm optimization (PSO) algorithm and Bessel transition curve. The simulation indicates that compared with traditional planning methods, the designed route is significantly smoother, and the total distance increase of the middle mobile is small. Li et al. [8] proposed a four-way search PP scheme suitable for mobile robots. This method achieves rapid optimization of PP by searching in four directions: horizontal and vertical. Compared with traditional

heuristic optimization algorithms, the four-way search scheme has strong advantages in solution space search, which helps to find high-quality paths with lower costs and satisfy various constraints. Yuan et al. [9] proposed a mobile PP algorithm for robots equipped with mobile sensor networks. Compared with traditional planning models, this algorithm has good adaptability and environmental awareness. The outcomes reveal that the algorithm in this study effectively improves the mobile PP ability of robots equipped with mobile sensor networks. The total length of the planned route is smaller than that of traditional planning models, and it has good collision avoidance ability. Liu and Jiang [10] proposed a PP model in view of the pigeon heuristic optimization algorithm. In the process of solving PP issues, this model takes the principle of avoiding collisions and unnecessary turns as much as possible to find a moving path. Through comparative experiments, it was found that this algorithm has superior PP performance in different scenarios. This indicates that its ability for finding the optimal path (OP) in complex environments exceeds other classical algorithms, providing an effective and efficient way to solve PP problems in complex scenarios. Meng et al. [11] proposed a safe and efficient LiDAR-based PP system to address the issue of insufficient obstacle avoidance ability in the PP of mobile robotic arms. This system is used to solve the navigation problem of four-wheel steering and four-wheel drive mobile robotic arms in manufacturing plants. In the study, the author utilized LiDAR technology to map the surrounding environment in real-time and identify obstacles. In addition, the study introduced a real-time collision avoidance algorithm to avoid dynamic and static obstacles. The simulation showcases that the PP system based on LiDAR proposed in the study exhibits high accuracy and robustness in handling navigation problems in manufacturing environments, and has excellent collision avoidance ability. Zhang et al. [12] presented a PP method for mobile robots in view of an improved local PSO algorithm. This algorithm increases the randomness and diversity in the local search process, improving the search capability. Through simulation experiments, the author verified the superior performance of this algorithm in mobile robot PP problems, indicating its good applicability and scalability in practical applications. This mobile robot PP method in view of improved local PSO algorithm provides an effective technical means to solve the navigation problem in complex environment in actual scenes.

In summary, although extensive research has been conducted to improve the planning performance and collision avoidance ability of intelligent PP models, there is little involvement in the construction of models that adjust adaptive parameters according to environmental characteristics. This mode is meaningful for improving the planning performance of PP models and the obstacle avoidance ability of robots.

III. ROBOT PATH PLANNING AND COLLISION AVOIDANCE STRATEGY IN VIEW OF IMPROVED ACO AND A* ALGORITHMS

The ACO algorithm was invented by imitating the foraging process of ants in nature, and has merits like strong adaptability and ease of utilizing in conjunction with other algorithms [13, 14]. For this reason, this algorithm was chosen in this study to construct a robot motion trajectory planning model. However, the ACO algorithm also has the disadvantage of being easily

trapped in local optima, so it is also essential for enhancing the algorithm. In view of the shortcomings of the ACO algorithm, it can improve its node selection, pheromone update mechanism, transition probability calculation method and volatility coefficient calculation method.

A. Design of Improved ACO Algorithm Based on Adaptive Parameter Setting

In actual working conditions, mobile robots will spend too much time in significant turning movements, and the corresponding energy loss will inevitably increase. Therefore, in PP, it is necessary to minimize significant turning points, improve the planning path, and diminish the motion cost of mobile robots. So now an improved planning path transition probability calculation method that integrates corner heuristic information is designed, as shown in Eq. (1).

$$P_{ij}^{(k)}(t) = \begin{cases} \frac{\tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)\omega_{ij}^\gamma(t)}{\sum_{s \in allowed_k} \tau_{ij}^\alpha(t)\eta_{ij}^\beta(t)\omega_{ij}^\gamma(t)}, & s \in allowed_k \\ 0, & otherwise \end{cases} \quad (1)$$

The superscript of the variable in Eq. (1) represents the ant number; The subscript is the path number; $\tau_{ij}^\alpha(t)$, $\eta_{ij}^\beta(t)$ and $\omega_{ij}^\gamma(t)$ respectively represent the pheromone quantity, heuristic information and corner heuristic function on the corresponding track of the corresponding ant at time t ; $allowed_k$ represents the feasible adjacency grid label set of ant k . The related calculating method of the corner heuristic function $\omega_{ij}(t)$ is depicted in Eq. (2).

$$\omega_{ij}(t) = \omega_i T \quad (2)$$

In Eq. (2), ω_i is an adjustable parameter with a value range of (0,1); T represents the turning cost. The corner of the path can be described as shown in Fig. 1, where the corners of paths (1) and (2) are both 45°, and the corners of paths (3) and (4) are 90° and 135°.

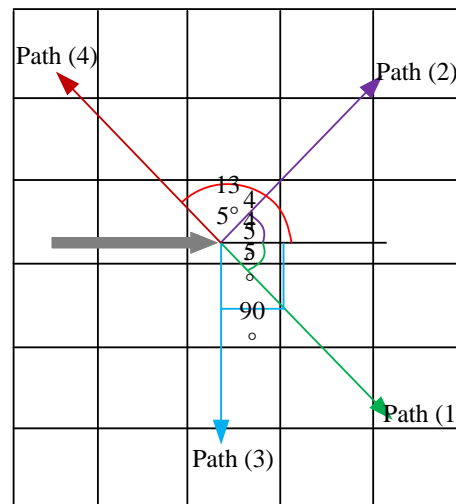


Fig. 1. Schematic diagram of improved path transition probability calculation method for path corners.

Therefore, T in Eq. (2) can be calculated according to Eq. (3).

$$\begin{cases} T = 2, \theta \leq 90^\circ \\ T = 1, 90^\circ \leq \theta \leq 180^\circ \end{cases} \quad (3)$$

In the traditional ACO algorithm, the path is selected according to the roulette wheel method, and there is a relatively significant pheromone concentration in the algorithm operation phase, which leads to the slow convergence of the ACO algorithm in the calculation premise [15, 16]. Meanwhile, in very large environmental conditions, traditional ACO algorithms may experience convergence stagnation, which significantly reduces the global solution optimization performance of the algorithm [17]. Therefore, an adaptive parameter setting method is designed here, which uses a mixture of random sexual selection and deterministic selection to calculate the selected path, as shown in Formula (4).

$$s = \begin{cases} \arg \left(\max \left(\left[\tau_{ij}(t) \right]^\alpha \cdot \left[\eta_{ij}(t) \right]^\beta \cdot \left[\omega_{ij}(t) \right]^\gamma \right) \right), q \leq q_0 \\ p_{ij}^k, \text{ else} \end{cases} \quad (4)$$

In Eq. (4) s represents the selected path; p_{ij}^k is the possibility of the corresponding ant appearing on the corresponding path; α , β and γ represent the corresponding ant numbers for $\tau_{ij}(t)$, $\eta_{ij}(t)$, and $\omega_{ij}(t)$, respectively; q is a uniformly distributed random variable in the range of [0,1], and q_0 is the threshold corresponding to the deterministic selection. The calculation method for q_0 is shown in Eq. (5). In this new adaptive parameter setting method, random variable q and deterministic selection threshold q_0 are used to show the randomness and certainty in the selection process respectively.

$$q_0 = \varepsilon \cdot \left\{ 0.2 + \left[\frac{N_{\max} - N_c}{N_{\max}} \right] \times 0.7 \right\} \quad (5)$$

In Eq. (5), ε is the adjustment coefficient; N_{\max} , N_c serves as the maximum quantity of iterations and the current quantity of iterations. Combining the search characteristics of ant colony during the iteration process, the initial q_0 of the improved ACO algorithm is generally greater than q , which means that the initial path of the improved ACO algorithm is determined in a pseudo random probability manner. In the later stage, as the small value of q_0 , ants are more likely to conduct random searches. It indicates that this strategy of selecting path nodes based on adaptive parameter calculation can effectively reduce the algorithm runtime, accelerate the algorithm convergence, and reduce the probability of stagnation. This increases the likelihood of the algorithm finding the optimal solution.

The traditional ACO algorithm will update the pheromone according to the way of updating all paths, but the disadvantage of this way is that the pheromone quantity of all paths may not differ greatly. This cannot highlight the competitiveness of the dominant path, and the convergence speed is slow. Therefore, to strengthen the attraction of the

path, the pheromone updating method is now adjusted. The calculation method of pheromone $\tau_{ij}(t+n)$ at time $t+n$ is shown in Eq. (6).

$$\tau_{ij}(t+n) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij}(t,t+n) + \frac{h(L_a-L_n)}{L_n} \quad (6)$$

In Eq. (6), L_a and L_n are the optimal values in the iteration history and the optimal values in this iteration, respectively; L_n is the adjustable coefficient. Therefore, when the calculation of each iteration of the algorithm is completed, when $L_a > L_n$, the corresponding path of this iteration is shorter. Eq. (6) will strengthen the pheromone strength of this iteration and save the OP generated in this iteration. On the contrary, if $L_a < L_n$, it means that the current path is not the shortest path. Eq. (6) will reduce the strength of pheromone.

Due to special terrain conditions, the constant volatility coefficient ρ will diminish the likelihood of the algorithm finding the OP. When the value of ρ is too large, although the algorithm converges faster, it also may fall into local optima; When the value of ρ is too small, the likelihood of previously explored nodes being repeatedly explored will increase, and the convergence speed will also decrease. Therefore, it now adjusts the size of ρ and obtains the value according to Eq. (7).

$$\rho = \begin{cases} 0.9\rho(t-1), L_a - L_n > b \\ 0.8\rho(t-1), a < L_a - L_n < b \\ 0.7\rho(t-1), L_a - L_n < a \end{cases} \quad (7)$$

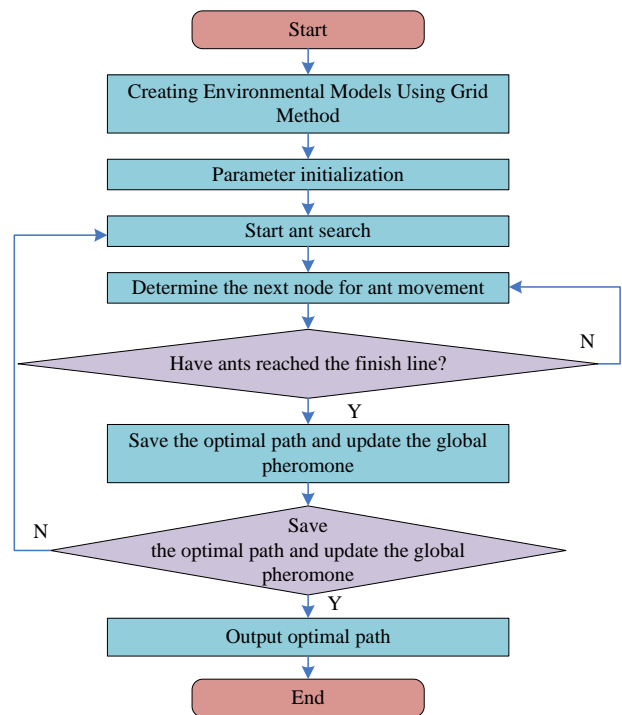


Fig. 2. Improved ACO algorithm calculation process.

In Eq. (7), L_A is the worst case path in the current calculation result; a and b are two constants; The smaller the $L_A - L_n$ value, the greater the likelihood of it falling into local optima, and the faster the update speed of ρ . The improved ACO algorithm used in robot movement PP has been designed, and its calculation is shown in Fig. 2.

B. Improved A* Algorithm and Robot Path Planning Model Design

Due to the fact that mobile robots often encounter moving obstacles in actual working environments, if the robot still follows the planned route, the time and energy consumption to reach the endpoint (EP) will increase, and even safety accidents may occur [18, 19]. Therefore, when designing a robot motion PP model, it is also necessary to consider obstacle avoidance issues. Robot movement PP cannot only consider the search efficiency of algorithms, but also the degree of twists and turns in the route. The classic A* algorithm is often applied to solve such problems, but it uses Manhattan distance and Euclidean distance to set heuristic functions, with only four search directions [20]. The fewer search directions in the classic A* algorithm will increase the number of probe nodes and the viewpoint of the path, which will affect the performance of the algorithm. Moreover, robots moving along winding paths can

also waste too much time. Consequently, it is essential for enhancing the traditional A* algorithm by performing path search calculations in eight directions, and combining the Manhattan distance $h_M(n)$ and Euclidean distance $h_g(n)$ of the current point n to design an estimated cost $h(n)$ that can simultaneously reduce the number of search nodes and bends. The calculation is demonstrated in Eq. (8).

$$h(n) = \max\left(\text{abs}(n_x - g_x), \text{abs}(n_y - g_y)\right) \quad (8)$$

In Eq. (8), $\text{abs}(\cdot)$ represents the absolute value operation; g_x and g_y represents the coordinates of the target node g_y in both axis directions. Therefore, the improved A* algorithm calculation process is demonstrated in Fig. 3.

In this study, an environment model for robot motion is established, as shown in Fig. 4. The basic method for modeling is the grid method, which selects grids of appropriate size to simulate the environment. The static fault objects in the environment in Fig. 4 are simulated using a blue grid, with a green cross representing dynamic obstacles (DO), and static obstacles that suddenly enter the environment are described in red. The white grid in Fig. 4 shows the areas where the robot can move freely.

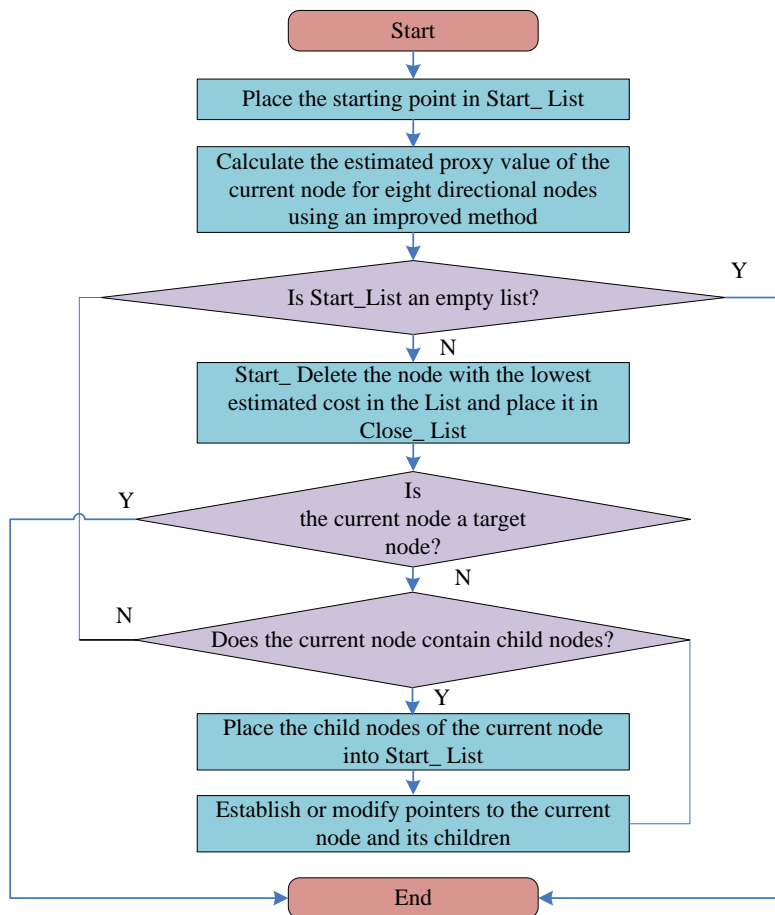


Fig. 3. Improved A* algorithm calculation process.

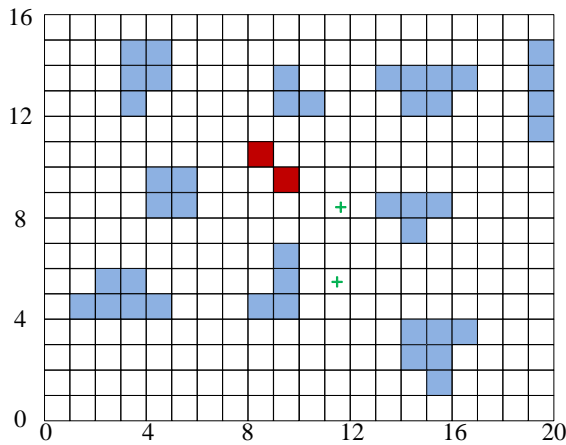


Fig. 4. Grid map of robot mobile environment simulation.

Robots need to continuously search for various obstacles within a limited range while moving, in order to make targeted obstacle avoidance behaviors. To perceive the movement path of the future DO, linear prediction models are now used to calculate the motion status of the DO. The current assumption is that the moving obstacle in the environment is $P(x, y)$, and the horizontal and vertical coordinate values in the orientation are correlated with time t . Therefore, (x, y) can be calculated according to Eq. (9),

$$\begin{cases} x = at + b \\ y = ct + d \end{cases} \quad (9)$$

g_y in Eq. (8) are unknown equation parameters. According to Eq. (9), calculate the corresponding (x_l, y_l) (where $l = 1, 2, \dots, n$) at n time points to obtain a linear equation in matrix form, as shown in Eq. (10).

$$\begin{cases} X_n = T_n + A_m \\ Y_n = T_n + A_m \end{cases} \quad (10)$$

The matrix content in Eq. (10) is shown in Eq. (11).

$$\begin{cases} X_n = [x_1, x_2, \dots, x_n]^T \\ T_n = \begin{bmatrix} t_1 & t_2 & \dots & t_n \\ 1 & 1 & \dots & 1 \end{bmatrix}^T \\ A_m = [a, b]^T \\ B_m = [c, d]^T \\ Y_n = [y_1, y_2, \dots, y_n]^T \end{cases} \quad (11)$$

In this study, the error vector E_n is set and calculated according to Eq. (12).

$$E_n = [e_1, e_2, \dots, e_n]^T \quad (12)$$

Therefore, Eq. (13) holds.

$$E_n = X_n - T_n \cdot A_n \quad (13)$$

In Eq. (13), A_n is the estimated value. The analytical error can be calculated according to Eq. (14).

$$J_n = \sum_{l=1}^n \lambda^{n-l} e_l^2 \quad (14)$$

The range of λ values in Eq. (14) is (0,1). Its redefinition F is calculated according to Eq. (15).

$$P_{n-1} = (T_{n-1}^T T_{n-1})^{-1} / \lambda \quad (15)$$

According to the observation values of moving obstacles at n different times, their spatial coordinates can be calculated, thereby calculating the specific values of parameters g_y . The behavior of the robot sensors, which detect the position of the DO in the circumstance will accompany the entire movement of the robot, used to estimate parameters for corresponding updates. Finally, the subsequent position information of the obstacle can be obtained according to the above method.

In real-world application scenarios, not all obstacles present in the environment of mobile robots are stationary, and there may be obstacles that can move, such as humans. Therefore, robots also need to avoid DO, which requires higher control and PP capabilities. Below is the design of collision prediction and obstacle avoidance strategies for robots. Considering the sensor measurement capabilities and movement methods of most mobile robots on the market, it is assumed that robots use their own sensors to continuously measure the movement position, direction, and speed of surrounding obstacles. The motion directions of robots and DO are described in Fig. 5.

Based on Fig. 5, the collision avoidance strategy designed in this study is illustrated: if there is no method of collision between the robot and the DO, the motion trajectory of the two needs to be observed to determine if there is an intersection between the two. If there are no intersections, it is assumed that they will not collide. In this case, the robot does not need to take additional collision avoidance actions and can move according to the originally planned trajectory, as shown in Fig. 5 for A, B, and F. But in the case where there is an intersection point between the robot and the obstacle's motion trajectory, the two will collide, and at this point, it is necessary to redefine the robot's motion path. For example, in the C and D motion situations in Fig. 5, the obstacle intersects with the robot's motion at the side, and the corresponding collision avoidance strategy is for the robot to remain stationary. After the current motion trajectory of the DO does not intersect with the robot, the robot continues to move along the set path. In summary, to better address the problem of robot motion trajectory planning, it is necessary to combine the improved A* algorithm with the improved ACO algorithm. The corresponding computation is shown in Fig. 6.

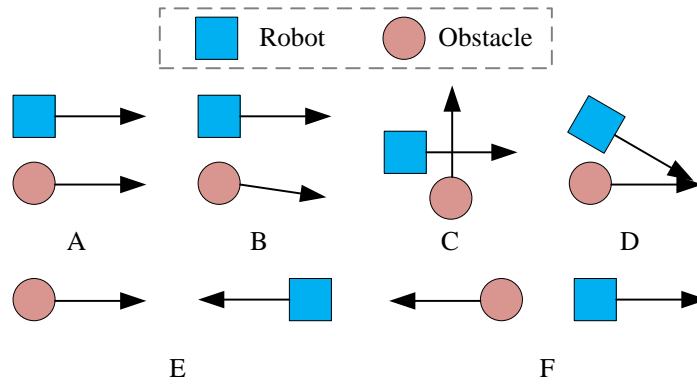


Fig. 5. Display diagram of dynamic obstacles and robot movement direction.

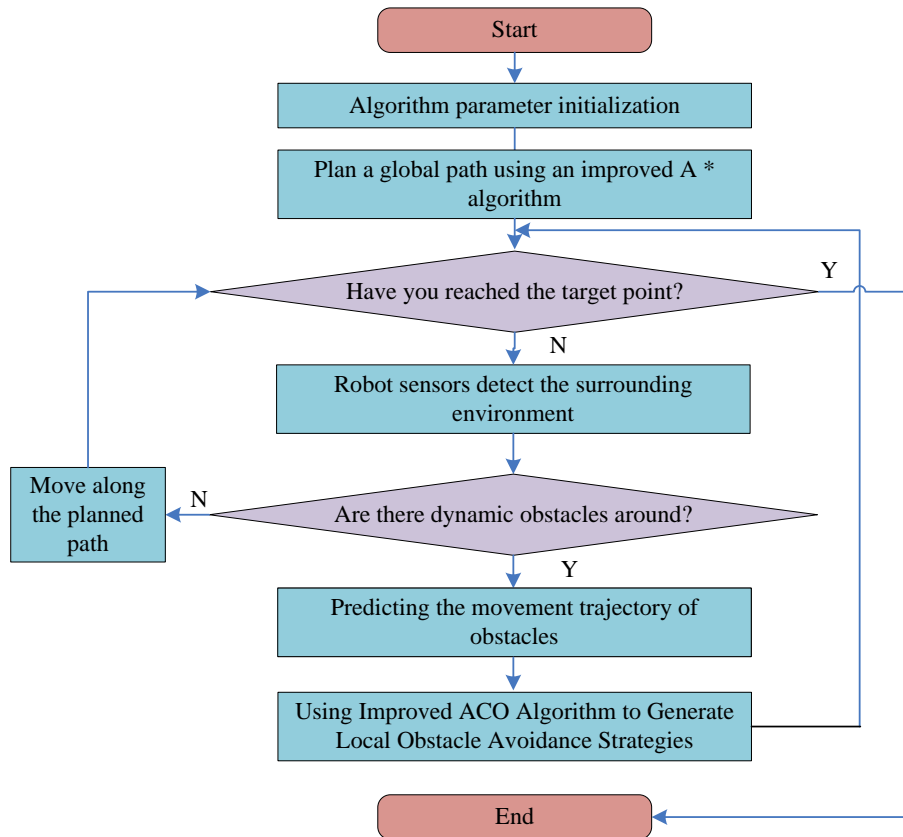


Fig. 6. Calculation process of robot path planning model using hybrid improved A* algorithm and IACO-A* algorithm.

IV. TESTING AND ANALYSIS OF ROBOT PATH PLANNING MODEL BASED ON HYBRID IMPROVED ACO ALGORITHM

After introducing a robot motion PP model in view of the improved ACO algorithm and the improved ACO-A* (IACO-A*) algorithm, the application of the model needs to be tested. In this study, only simulation experiments are used to verify the design model. Meanwhile, to simplify the experiment, the mobile environment in the test is abstraction into a grid map. The simulation experiment was run on the MATLAB2016 platform, and the parameters of the IACO-A* algorithm were determined through multiple trial runs as follows: the maximum number of cycles was 50, $\alpha = 4$, $\beta = 8$, $\rho = 0.7$,

and the total quantity of ants was 50. The experimental circumstance is a two-dimensional map, and the minimum size of the tested grid is 10 m × 10 m, with a maximum size of 100m × 100m, with a grid growth step of 5m. In the experiment, the widely used ACA, PSO algorithm, and the novel Beetle Antennae Search (BAS) algorithm were selected to construct a comparative planning model. The parameters of the comparative model were also determined according to the trial operation method. The robot is set to move forward at a constant speed of 1m/s when no DO is encountered. After encountering obstacles and temporarily stationary, it first accelerates to 1m/s at an acceleration of 1m/s², and then continues to move at a constant speed. There are two DO in the

grid map, and their initial positions appear randomly. Referring to the actual working circumstance of mobile robots, the proportion of static obstacle grids to the total quantity of grids should be within the range of 20% to 70%; within this numerical range, the quantity of static obstacles generated is also random, and the setting results are shown in Table I.

TABLE I. COMPARISON OF ALGORITHM PARAMETER SETTING SCHEMES

Number	Algorithm name	Parameter Name	Numerical value
#1	ACO	Maximum number of cycles	50
		Ant number	200
		Walking distance	1.5
		Pheromone volatilization factor	0.6
#2	BAS	Maximum number of iterations	50
		The initial length of the antennae of the longicorn beetle	10
		Step decay factor	0.95
		Maximum step size	1.26
		Minimum step size	0.35
#3	PSO	Maximum number of iterations	50
		Inertia weight	0.6
		Learning factor	1.4
		Maximum speed	27
		Minimum speed	4

Firstly, a specific PP analysis is carried out using the planning results with a minimum grid size of 10 * 10 m and no DO conditions as a representative. The grid map generation results are randomly selected, and the OP planned by IACO-A * and ACO algorithms is indicated in Fig. 7. The horizontal and vertical axes in Fig. 7 represent the X and Y axes in the two-dimensional grid map, respectively; The scale units are all in meters; The blue grid represents the static obstacles that exist at the initial moment; The green dots and red crosses represent the starting point (SP) and EP of the robot's movement path, respectively; The black dashed line serves as the planned movement path. Fig. 7 demonstrates that the total path lengths of IACO-A * and ACO algorithms are 7.95 m and 12.64 m, respectively. Both the enhanced and the improved ACO algorithms can enable the robot to move from the SP to the EP; But the route planned by the IACO-A * algorithm has significantly fewer bends, and the overall route is smoother, resulting in a shorter motion time.

Further analysis of the planning results was conducted using 10 * 10 m DO conditions. Fig. 8(a), 8(b), 8(c), and 8(d) represent the initial planning path, the path when DO, O1 and O2 are added, the path to avoid O1, and the path to avoid O2, respectively. The green color in Fig. 8 serves as the path of the robot; Black serves as the movement path of DO; the dashed line serves as the planned path or the path that has been taken; The dotted line serves as the path of the obstacle or the robot during the avoidance process. Fig. 8(b) and 8(c) indicate that the robot will encounter side collisions with O1 DO during its movement. Therefore, choose to stay in place for a period of time until it is calculated that there is no longer an intersection between the two routes before continuing to move along the

original route. Observing Fig. 8(d), it can be seen that after the robot detects a frontal collision with the DO O2, the IACO-A * algorithm generates local target points that modify the original path to some extent. After avoiding O2, the robot moves along the original path.

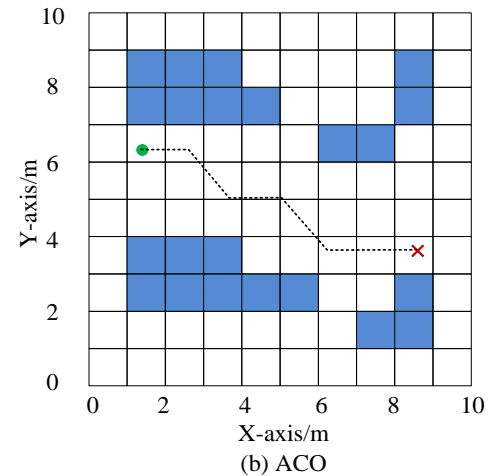
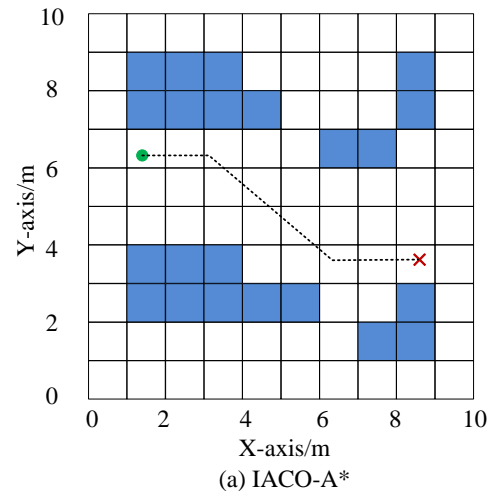
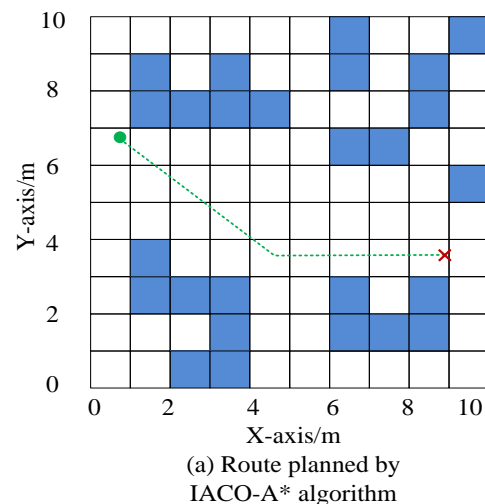
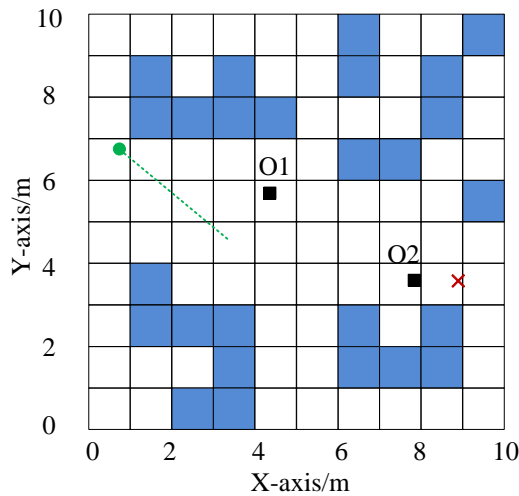
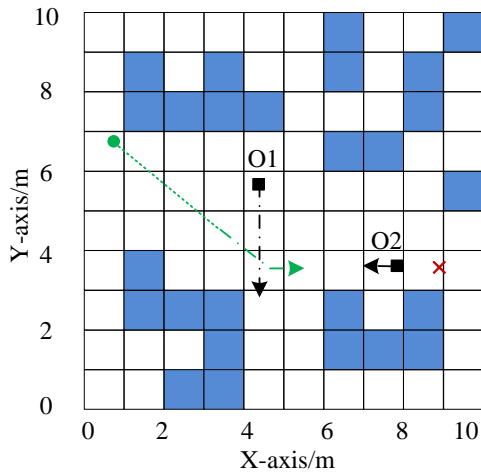


Fig. 7. IACO-A * and ACO algorithm in 10 × 10 m optimal planning path in a 10 m grid map.

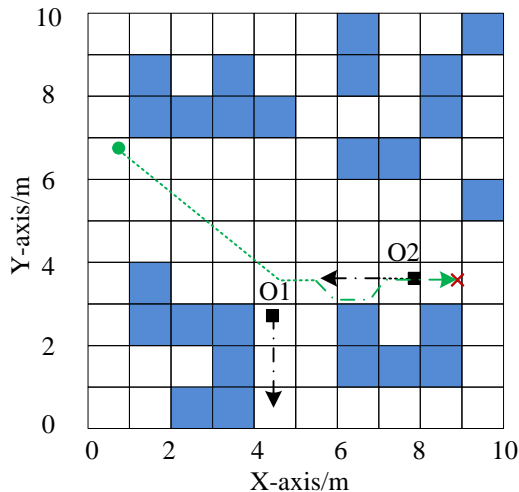




(b) Environment when adding dynamic obstacles



(c) Path to avoid dynamic obstacle 1



(d) Path to avoid dynamic obstacle 2

Fig. 8. IACO-A * in 10 with dynamic obstacles × planned path in a 10m map.

It reanalyzes the planning outcomes of the unimproved ACO algorithm in the presence of DO, and the results are shown in Fig. 9. The meanings of the horizontal and vertical axes, icons, and lines in Fig. 9 are consistent with those in Fig. 8. Fig. 9 shows that although the route planned using the traditional ACO algorithm also avoids DO O1 and O2, the adjusted moving route significantly detours compared to Fig. 8. Based on Fig. 8 and 9, it is found that using the IACO-A * algorithm and the ACO algorithm, the overall travel distance of the two algorithms is 12.28 m and 16.74 m, respectively. This indicates that the IACO-A * algorithm exceeds the traditional ACO algorithm in the overall obstacle avoidance capability and route rationality of the robot PP.

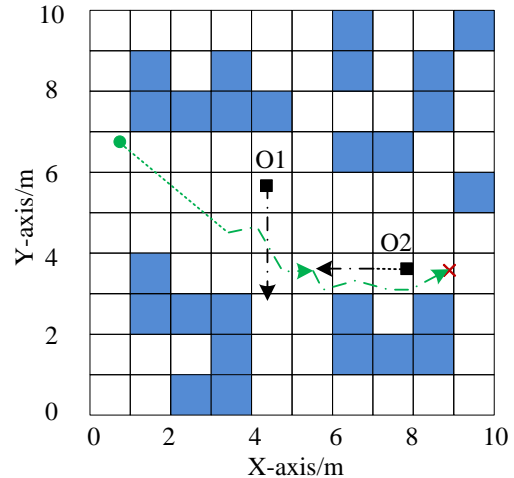
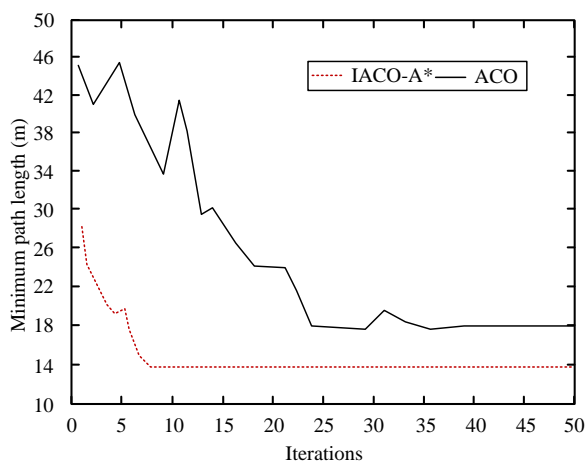


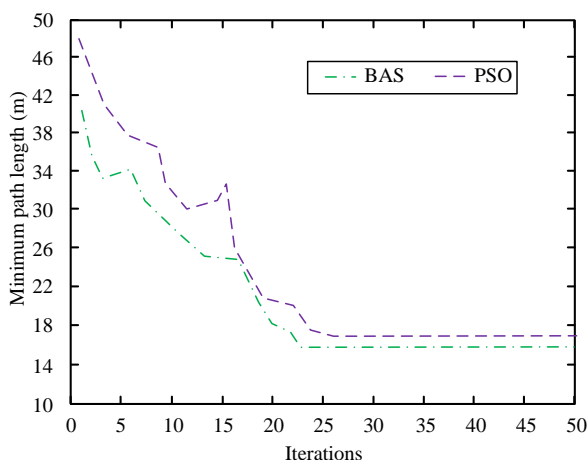
Fig. 9. ACO in 10 with dynamic obstacles × planned path in a 10 m map.

Next, it selects all the comparison models and calculates their minimum path length changes with the IACO-A * planning model during the training process of a 10 * 10m grid map with DO. It is indicated in Fig. 10. In Fig. 10, the horizontal axis (HA) serves as the quantity of iterations, while the vertical axis represents the minimum path length planned, in meters. Different subgraphs and line styles represent different planning models. Fig. 10 shows that the IACO-A *, ACO, BAS, and PSO planning models converge after 8, 37, 23, and 26 iterations, respectively. The minimum path lengths after convergence are 13.24m, 17.82m, 16.24m, and 17.05m, respectively. The experiment indicates that the PP model in this study based on the IACO-A * algorithm has the fastest convergence speed during training, and the total length of the OP after convergence is the smallest.

The performance of each planning model in the test dataset of a 10 * 10m grid map with DO is demonstrated in Fig. 11. The HA in Fig. 11 serves as repeated experimental tests, which are conducted to verify the stability of the output results of each planning model. Fig. 11 shows that when the number of repetitions is small, the minimum length fluctuation of the output routes of each planning model is more severe. However, as the number of repetitions increases, the minimum path length fluctuation gradually decreases. The minimum planning length standard deviations for IACO-A *, ACO, BAS, and PSO planning models are 0.82m, 1.24m, 0.95m, and 1.78m, respectively.



(a) IACO-A * and ACO



(b) PSO and BAS

Fig. 10. Minimum path length of each planning model in training.

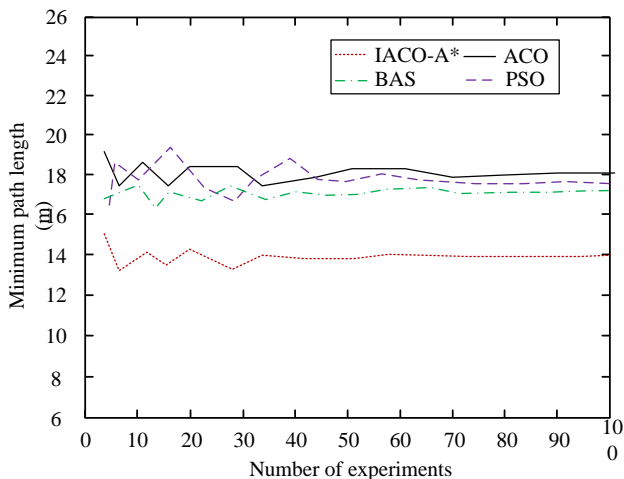
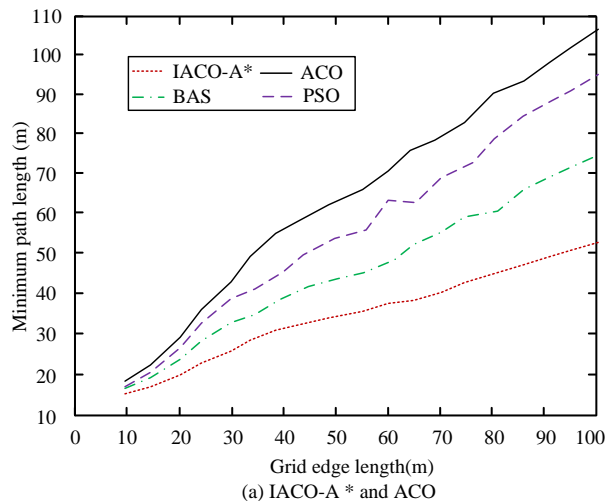


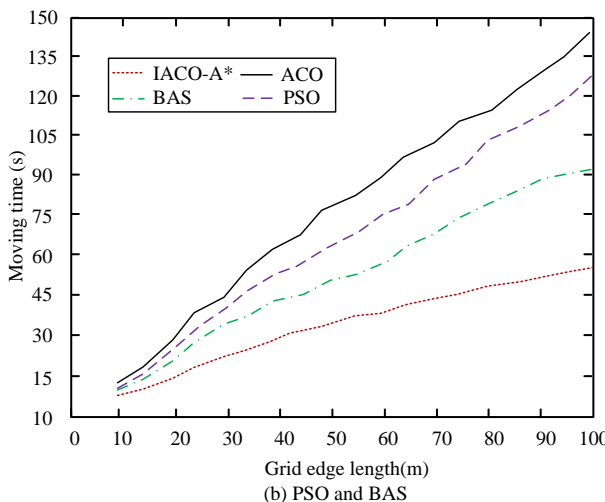
Fig. 11. Each planning model with dynamic obstacles 10 × Performance in 10 m grid map testing.

It further analyzes the minimum planning path length and corresponding movement time of each planning model under different grid size maps with DO. It is illustrated in Fig. 12. The HA in Fig. 12 serves as the edge lengths of different grid maps, in meters; The vertical axes of Fig. 12(a) and 12(b)

represent the minimum travel path length and the corresponding total travel time, in units of m and s, respectively. The line style is used to distinguish the planning model. Fig. 12 serves as that as the size of the map grid grows, the time and total distance required for the robots to move from the SP to the EP in the map also increase. However, regardless of the size of the grid map, the minimum path length and total movement time of the IACO-A * planning model designed in this study are smaller than all the comparative planning models. For example, when the edge length of the grid map is 100m, the minimum planning length and total movement time of IACO-A *, ACO, BAS, and PSO planning models are 49m, 104m, 75m, 93m and 49s, 142s, 93s, and 127s, respectively.



(a) IACO-A * and ACO



(b) PSO and BAS

Fig. 12. Minimum path length and movement time of each model under different map sizes.

In order to further improve the reliability of the research results, the IACO-A * model is now placed in a domestic mobile robot product to carry out dynamic obstacle avoidance experiments in planar and curved scenes. The dynamic obstacles are spheres, rectangular parallelepipeds, and regular tetrahedrons, respectively. After repeated experiments, it was found that the robot system equipped with the IACO-A * model has the fastest dynamic obstacle avoidance response speed and the shortest obstacle avoidance time. Its practical

performance is significantly better than that of mobile robots equipped with other models.

V. RESULTS AND DISCUSSION

This study designed an improved ACO path planning model based on adaptive parameter setting, and conducted dynamic obstacle avoidance path planning experiments in a simulation grid map. The experimental results show that the model designed in this study has a grid size of $10 \times$ the total planned path length under the conditions of 10 m and no dynamic obstacles is 7.95 m, which is lower than the output path length of the comparative model and can reach the set focus normally. Simultaneously, in the case of a grid size of $10 * 10$ m and the presence of dynamic obstacles, the IACO-A * planning model corresponds to the robot using pause and generate local target points to avoid obstacles O1 and O2 that may cause side and frontal collisions, respectively. The ACO algorithm corresponds to the robot avoiding O1 and O2 obstacles with the same movement mode by generating local target points, but the overall movement distance is 16.74 meters, which is significantly higher than the data of the IACO-A * planning model. Moreover, in various scenarios where the grid edge length of the simulated grid map changes from 10 to 100, the total path length of the planning model output in this study is consistently lower than that of all the comparative models, and also lower than the total path length under the same conditions as in references [9] and [11]. Finally, the IACO-A * model was placed in a domestic mobile robot product to carry out dynamic obstacle avoidance experiments in planar and curved scenes. The dynamic obstacles have the shapes of sphere, rectangular cuboid, and regular tetrahedron, respectively. After repeated experiments, it was found that the robot system equipped with the IACO-A * model has the fastest dynamic obstacle avoidance response speed and the shortest obstacle avoidance time. Its practical performance is significantly better than that of mobile robots equipped with other models.

From the results of various types of experiments conducted in this study, it can be seen that the robot movement path planning model based on the improved ACO algorithm designed in this study has excellent path planning and obstacle avoidance capabilities, which can enable the robot to smoothly and efficiently avoid dynamic obstacles in the route.

VI. CONCLUSION

In this study, an improved ACO algorithm and an improved A * algorithm were developed, and the two were fused to construct a robot motion PP model. The experiment reveals that under the conditions of grid size $10 * 10$ m and no DO, the total path lengths of IACO-A * and ACO algorithms are 7.95m and 12.64m, respectively. Both the enhanced and the improved ACO algorithms can enable the robot to move from the SP to the EP. Under the conditions of a grid size of $10 * 10$ m and the presence of DO, the IACO-A * planning model corresponds to robots that use pause and generate local target points to avoid obstacles O1 and O2 that may cause side and frontal collisions, respectively. The ACO algorithm corresponds to the robot that avoids obstacles O1 and O2 with the same motion mode by generating local targets; however, the overall mobile distance of 16.74m is significantly higher

than the data of the IACO-A * planning model. During the training process, the IACO-A *, ACO, BAS, and PSO programming models completed convergence after 8, 37, 23, and 26 iterations, respectively. The minimum path lengths after convergence were 13.24m, 17.82m, 16.24m, and 17.05m, respectively. When the edge length of the grid map is 100m, the minimum planning length and total movement time of the IACO-A *, ACO, BAS, and PSO planning models are 49m, 104m, 75m, 93m and 49s, 142s, 93s, and 127s, respectively. The experimental data prove that the robot motion PP model designed in this study with respect to the improved ACO algorithm has excellent PP and obstacle avoidance capabilities. However, due to research limitations, it was not possible to invite multiple industry experts to subjectively evaluate the practicality of the model. From the results of this experiment, it can be seen that the shape of obstacles has an extremely small impact on the path planning results of the IACO-A * model.

ACKNOWLEDGMENT

The research was supported by 2020 Scientific and Technological Innovation Programs of Higher Education Institution in Shanxi: "Research and Development of Industrial Robot Trajectory Planning Algorithm Optimization and Experimental Device (No.: 2020L0760)".

REFERENCES

- [1] Y. Li, W. Wei, Y. Gao, D. Wang, and Z. Fan, "PQ-RRT*: An improved path planning algorithm for mobile robots," *Expert Syst. Appl.*, vol. 152, pp. 113425, April 2020.
- [2] Y. Liu, Y. Li, Y. Niu, and D. Jin, "Joint optimization of path planning and resource allocation in mobile edge computing," *IEEE Trans. Mob. Comput.*, vol. 19, no. 9, pp. 2129-2144, June 2020.
- [3] Q. Liu, L. Shi, L. Sun, J. Li, M. Ding, and F. S. Shu, "Path planning for UAV-mounted mobile edge computing with deep reinforcement learning," *IEEE Tran. Veh. Technol.*, vol. 69, no. 5, pp. 5723-5728, March 2020.
- [4] X. Zhang, J. Lai, D. Xu, H. Li, and M. Fu, "2D Lidar-based SLAM and path planning for indoor rescue using mobile robots," *J. Adv. Trans.*, vol. 2020, no. Pt.9, pp. 8867937, November 2020.
- [5] Y. H. Hsu and R. H. Gau, "Reinforcement learning-based collision avoidance and optimal trajectory planning in UAV communication networks," *IEEE Trans. Mob. Comput.*, vol. 21, no. 1, pp. 306-320, June 2020.
- [6] F. O. Coelho, M. F. Pinto, and J. Souza, "Hybrid methodology for path planning and computational vision applied to autonomous mission: A new approach," *Robotica*, vol. 38, no. 6, pp. 1000-1018, July 2020.
- [7] L. Xu, M. Cao, and B. Song, "A new approach to smooth path planning of mobile robot based on quartic Bezier transition curve and improved PSO algorithm," *Neurocomputing*, vol. 473, no. Feb.7, pp. 98-106, February 2022.
- [8] K. Li, C. Yuan, J. Wang, and X. Dong, "Four-direction search scheme of path planning for mobile agents," *Robotica*, vol. 38, no. 3, pp. 531-540, 2020.
- [9] W. Yuan, N. Ganganath, C. T. Cheng, G. Qing, C. M. Lau, and Y. Zhao, "Path-planning-enabled semiflocking control for multitarget monitoring in mobile sensor networks," *IEEE Trans. Ind Inform.*, vol. 16, no. 7, pp. 4778-4787, July 2020.
- [10] A. Liu and J. Jiang, "Solving path planning problem based on logistic beetle algorithm search-pigeon-inspired optimisation algorithm," *Electron Lett.*, vol. 56, no. 21, pp. 1105-1108, 2020.
- [11] J. Meng, S. Wang, Y. Xie, G. Li, X. L. Zhang, L. Jiang, and C. Liu, "Safe and efficient navigation system for 4WS4WD mobile manipulator in manufacturing plants," *Meas. Sci. Technol.*, vol. 32, no. 4, pp. 45203, November 2020.

- [12] L. Zhang, Y. Zhang, and Y. Li, "Mobile robot path planning based on improved localized particle swarm optimization," *IEEE Sens J*, vol. 21, no. 5, pp. 6962-6972, November 2020.
- [13] M. Wahab, S. Nefti-Meziani, and A. Atyabi, "A comparative review on mobile robot path planning: Classical or meta-heuristic methods?" *Annu. Rev. Control*, no. 50, pp. 233-252, October 2020.
- [14] X. Deng, R. Li, L. Zhao, K. Wang, and X. Gui, "Multi-obstacle path planning and optimization for mobile robot," *Expert Syst. Appl*, vol. 183, pp. 115445, November 2021.
- [15] J. Zan, "Research on robot path perception and optimization technology based on whale optimization algorithm," *J. Comput. Cognitive Eng*, vol. 1, no. 4, pp. 201-208, March 2022.
- [16] S. Wu, Y. Du, and Y. Zhang, "Mobile robot path planning based on a generalized wavefront algorithm," *Math. Probl. Eng*, vol. 2020, no. Pt.6, pp. 6798798, March 2020.
- [17] L. Tiseni, D. Chiaradia, M. Gabardi, M. Solazzi, D. Leonardis, and A. Frisoli, "UV-C mobile robots with optimized path planning: Algorithm design and on-field measurements to improve surface disinfection against SARS-CoV-2," *IEEE Robot. Autom. Mag*, vol. 28, no. 1, pp. 59-70, January 2021.
- [18] T. Hossain, H. Habibullah, R. Islam, and R. V. Padilla, "Local path planning for autonomous mobile robots by integrating modified dynamic-window approach and improved follow the gap method," *J. Field Robot*, vol. 39, no. 4, pp. 371-386, December 2021.
- [19] J. Li, J. Sun, L. Liu and J. Xu, "Model predictive control for the tracking of autonomous mobile robot combined with a local path planning," *Meas. Control*, vol. 54, no. 9/10, pp. 1319-1325, October 2021.
- [20] W. Chi, Z. Ding, J. Wang, G. Chen, and L. Sun, "A generalized voronoi diagram based efficient heuristic path planning method for RRTs in mobile robots," *IEEE Trans. Industrial Electron*, vol. 69, no. 5, pp. 4926-4937, May 2021.

Simulation of Logistics Frequent Path Data Mining Based on Statistical Density

Fengju Hou*

Zibo Vocational Institute, Shandong, China

Abstract—Sharp increases and rapid development followed the effects of a novel coronavirus outbreak on online sales and the real economy. The e-commerce mode on the Internet has attracted much attention, and users' purchases on the Internet has never been done before. However, among the many express companies, as the ones closest to consumers, they can still provide high-quality products in the face of huge market demand. Urban terminal logistics refers to the purpose of express services to meet the needs of terminal customers under the requirements of logistics centralization and customer diversification. However, the geographical distribution of logistics services in China is comprehensive, and customers' requirements are also complex. Practical problems in logistics enterprises in China significantly restrict the quality of logistics services. The final kilometer of distribution is composed of many links, and it is a very cumbersome enlance; it contains the determination of distribution scope, loading goods, arrangement of distribution sequence, arrangement of vehicles or personnel scheduling, and planning of distribution routes. A Genetic Algorithm (GA) with local search method fusion is proposed for fast logistics data modeling and mining simulation analysis. Practical examples and literature data prove the method's accuracy.

Keywords—Statistical density; logistics; the path; the simulation data

I. INTRODUCTION

Since transportation costs have such a significant impact on the whole process, it is essential to optimize coordination routes in order to minimize transportation time and costs. This method increases efficiency while decreasing overall transportation costs. From the perspective of transportation constraints, Cheng Nan constructed a terminal transportation route containing transportation constraints in order to reduce transportation costs. Li Kunyenga proposed a basic mathematical model based on linearization by linking logistics' terminal characteristics with customers' needs. Based on this model, he proposed a new branch and truncation method, which included a new method to decompose the efficiency inequalities. The study [1] proposes using vehicle routing planning, joint distribution, and various forms to optimize rural distribution with low-density, long lines. The Genetic Algorithm (GA) is applied to optimize "at the end of the line" and maximize vehicle capacity and distribution costs. Using quadratic programming, [2] proposed a logistics distribution model that optimizes warehouse-to-distribution-center direction, customers, and express transport routes. Implementing a multigroup GA, its effectiveness was validated and confirmed. Urban terminal logistics distribution

in China still has these issues due to complex and changing work.

The cost of transportation remains exceptionally high. For now, logistics costs remain high among end-city logistics companies. The current problems in China are as follows: In the tip area of China, because the customer group is widely distributed, it needs a lot of human and material resources investment in infrastructure, equipment, and other aspects to increase investment, such as the establishment of the automatic warehouse and intelligent automatic suitcase, not only to invest heavily but also to increase the cost of project operation and maintenance [3–4]. Ineffective logistics is due to a lack of resources because the requirements of each user for logistics are very different, in the whole logistics system; no one can realize the coordination process, which reduces the efficiency of the whole logistics system. There is a delay in the delivery of the goods. In the end, the logistics, according to their actual situation, were to conduct the distribution. Our country has many subjective reasons, such as imperfect traffic routes. In addition, to some areas, the scale of logistics enterprises is small, and they often need to stay in the distribution center for a short time before distribution, which leads to an extended delivery time. Most express companies have not developed a set of perfect reward and punishment mechanisms in the process of terminal distribution, which leads to problems of lax work attitudes among their staff, product loss, damage, and delayed delivery, resulting in low customer satisfaction [5–6].

According to the current research results, various intelligent algorithms are applied to logistics path planning to achieve efficient and intelligent path optimization. It is the most commonly used path planning method. With the increasing intelligence of the express delivery industry in China, many people have begun to discuss how to establish an efficient online delivery system from the perspectives of intelligent self-lifting boxes, distribution center location, and distribution path selection. Zhang Shabo put forward the cost model, pricing model, and robust model according to the characteristics of intelligent express, analyzed the case with some examples, and put forward some reference opinions on how to construct a terminal distribution network in the construction and operation. Wu Maupye et al. analyzed the key problems faced by terminal logistics in property-free communities from two perspectives of the introduction and investment of express boxes, as well as optimization strategies such as terminal logistics collaborative distribution, standardized logistics management in property-free communities, optimized pickup resource allocation, and

*Corresponding Author.

omnichannel inventory operation management in order to provide decision-making reference for the intelligent express cabinet to solve the "last kilometer" terminal logistics in the non-realty community. Lee's action is based on data mining, to solve the problem of terminal distribution mode selection, through this technology, build a terminal distribution mode selection, and solve the data instance, finally using cluster algorithm to solve the problem of the terminal distribution system, through the case study, Finally, a better result than the existing distribution system is obtained according to the "delivery to home" and "customer pick up" characteristics.

On the choice of the distribution center, Yin Xiaoqing 8 and others, based on travel time reliability and a complex network structure, built a city distribution station at the end of the cold chain planning model and conducted the simulation calculation. An example that confirms the feasibility and effectiveness of the method is the assurance of utmost timeliness. Hi-ping Ren proposed the establishment of a network in the midst of the sorting center with end customers, with the aim of resolving the intricate issue of "end" distribution for retail enterprises. The network functions as a link between the customer and the sorting center, arranging the distribution sequence from the distribution center to the client's terminal within the classification and distribution network. Terminals could be office buildings, residential areas, or districts. The existence of a hierarchical sorting mode makes the business between employees and customers less, minimizing the consumption of time, but also reduces the logistics company to the user's delivery demand and the overall layout of the delivery mode, so that in the case of meeting consumer needs more actively for consumers and the company to bring a win-win situation. Zhao Xuedong for all kinds of fresh food, all kinds of logistics center layout and route optimization, set up a balanced carbon dioxide emissions and customers at the lowest transport demand stratified programming model, and under different fresh demand, the selection of logistics center and transportation route setting has a significant effect. An example has proved the correctness of this method. Based on providing efficient, safe, and fast distribution services for e-commerce customers, John Fernier et al. set up a regional distribution center to meet customers' high-quality service requirements and, on this basis, reduce logistics costs [7-9].

The paper is organized as follows: Section II presents the research method, Section III presents the proposed methods and result analysis of this work, and Section IV concludes the work.

II. RESEARCH METHOD

A. Overview and Application of Statistical Density Methods

In mathematics and technology, the most significant logistics problem is the optimization of statistical density. The frequent logistics routing problem is a hybrid optimization problem with various representative problems. As soon as the research results of statistical density are published, it has attracted the attention of many researchers. A passenger starts from one place, sells things in another, and returns to the origin. There is only one visit to a place. When the distance between the two places is known, he can decide his optimal

way of walking according to his situation. Before the logistics enterprise is discussed in detail, this article will elaborate on the basic knowledge of the logistics enterprise problem - the basic idea of the authorization map. In graph theory, the frequent problem of logistics routing is the minimum Hamiltonian cycle method, and the Hamiltonian cycle is a kind of node without a branch. A weighted graph is an undirected graph with a weight attached to each edge. Denoted as a weighted graph, it is a set of vertices, a set of edges, and represents the distance between vertices, Eq. (1)

(1)

Let denote the sub path from to, which takes the value 1 if it exists in the travelling salesman path and 0 otherwise, and the expression is given below, Eq. (2) and Eq. (3).

$$X_{ij}ij(i \rightarrow j)X_{ij} \quad (2)$$

$$X_{ij} = \begin{cases} 1, (i, j) \in L \\ 0, (i, j) \notin L \end{cases} \quad (3)$$

where is the solution sequence, then the objective function of the travelling salesman problem is expressed as a formula, and the travelling salesman's total distance through the path must be minimized, Eq. (4) and Eq. (5).

$$L(i, j) \in E \quad (4)$$

$$\begin{aligned} mZ &= \sum_{i \neq j} d_{i,j} X_{i,j} \\ mZ &= \sum_{i \neq j} d_{i,j} X_{i,j} \\ \sum_{i \in V, i \neq j; x_{ij}} &= 1 \\ \sum_{j \in V, j \neq i; x_{ji}} &= 1 \\ \sum_{i, j \in V; x_{ij}} &= |L|, L \in E \end{aligned} \quad (5)$$

It is a solution and then the objective function of the frequent logistics routing problem is expressed by the equation above, which requires the shortest total distance of the route through which the logistics need to pass.

In algorithmic terms, a publicly recognized data set tests and evaluates the computational power to solve a problem. It contains the best test data for a particular problem and the best solution for a problem [10-14].

B. Simulation of Logistics Frequent Path Data based on Statistical Density

This paper proposes a standard model based on statistical density simulation interval LIB for the frequent logistics path problem. The first statistical density simulation interval LIB found originates in some previous research articles by Dantzig, Fulkerson, Johnson Held and Karl, Karr, and Thompson et al. GerhardReinelt collected many practical problems, for example, in industry and geography, and gradually accumulated, eventually resulting in preliminary experimental data. LIB's statistical density simulation interval currently includes more than 100 examples, covering a wide range,

from 14 small to 85,900, strongly supporting solving logistics enterprise problems. However, some of these examples are fabricated purely to solve the frequent person problem of logistics routing, such as the statistical concentration modulus simulation value of sample 225, the correction of the LIB of the statistical concentration simulation has been stopped, and the results of a large number of statistical simulation experiments.

The statistical concentration simulation section LB includes three test data:

- 1) The symmetry of traveller's.
- 2) Hamiltonian ring problem.
- 3) Asymmetric logistics path frequent person problem.

On this basis, in the statistical density simulation region of LB, the format of the test data set that can be used to assess the method's performance is not uniform.

Location of the space: In this format, each point is identified with a number from 0 to the largest, and the coordinates of this point in the coordinate system are given according to the number. The initial test data is the spacing between all points in this format. In this format, the initial test data are the upper triangles of the spacing matrix between points. It is divided into diagonal, nondiagonal, row-first, and column-first, according to the order of precedence of whether there is a diagonal or not and row-first [15].

In the swarm, each person is an artificial intelligence that randomly and progressively constructs a solution to a class of optimal problems. Statistical density, a traditional algorithm, has been used more in road problems because its efficiency has always been the focus of attention, Eq. (6).

$$p_k(i, j) = q \left\{ \frac{[\tau(i, j)]^\alpha [\eta(i, i)]^\beta}{\sum_{\tau \in J_k(i)} [\tau(i, j)]^\alpha [\eta(i, i)]^\beta} \right\}; j \in J_k(i) \quad (6)$$

Then, as shown in the above formula, under statistical concentration will, calculate the migration of a city's risk, and then according to the "roulette gambling" method to identify the migration of the next city, in this process, the number of forbidden areas will be more, until every city covered by the taboo table, will stop for detection of the region.

Among these factors, the pheromone has the most significant coefficient of importance, and the route taken by the ants is randomized as the data changes. The lower the selected value is, the randomness of the stochastic optimization can be improved, but at the same time, the algorithm's convergence rate is reduced. As a key factor of the heuristic function, if its value is too large, it will be randomized so that it is easy to enter the optimal. In the case of abbreviated time, the randomization performance of the algorithm is improved, but the difficulty of the solution is also increased.

Then, the pheromone changes. After getting all the practical solutions, the global pheromone update is conducted according to the scheme. However, when searching, we adjust the signal to the pheromone change to obtain a new expression for the time series pheromone. Let us say that we want to

minimize the error of the objective function, expressed as Eq. (7).

$$K(C1, C2, \dots, Ck)ZZ \quad (7)$$

$$Z = \sum_{i=1}^k \sum_{x \in c_i} \|x - u_i\|^2$$

First, given the initial statistical density center, the algorithm will be iterated in the following two steps: a) Partition: The formula indicates that each node is divided into clusters such that the sum of squares within the cluster is minimized, Eq. (8)

$$m_1^{(0)}, m_2^{(0)}, \dots, m_k^{(0)} \quad (8)$$

The statistical density method is known as the average method. Its idea is quite simple. A sample set is divided into several small groups according to the distance of the sampling interval to ensure that the point groups are concentrated as far as possible and that the group is divided as clearly as possible. This paper proposes a method based on time series to solve the large-scale probabilistic simulation problem. It can divide an uneven urban area into several groups. The calculation accuracy is low, and the result is unstable when calculating the simulation area with less statistical density.

III. ANALYSIS OF RESULTS

A. Large-scale Statistical Density Analysis of Logistics Routing Data

After clustering in a large-scale statistical density simulation interval, the best route is determined based on the probability distribution of each simulated area. The pheromone is the key factor of the research object, and its change is related to the effect of the whole method. Under the traditional computational density, the starting concentration of each channel is the same and constant because the length of the channel will increase, so the subject will be attracted to the higher concentration so that the least one can be found in the screening repeatedly. However, this method has a drawback: it can lead to finding an incorrect route under uncertainty, resulting in a longer, time-consuming search.

Therefore, this paper presents a correction method based on point distance to correct the pheromone density of adjacent sides. We select a minimal path from a line with a start and end and multiple paths. In this line, the distance of the path is represented by the straight-line distance from the start point to the endpoint. The initial pheromone concentration between nodes can be defined as follows: Eq. (9) and Eq. (10)

$$d_{AC} D_{Aj} + d_{jc} A - j - C_{ij} \quad (9)$$

$$\tau_{ij}(0) = \frac{d_{AC}}{d_{Aj} + d_{jc}} \quad (10)$$

According to the property that pheromone is inversely proportional to path length, the above equations indicate that when the node is closer to the edge, that is, the closer it is to the shortest straight line distance, the value of is closer to 1,

and the statistical density will be biased towards such node movement.

This method guides the primary research target to avoid blind searches to improve the solution rate. After determining the new, improved method, the operation flow parameters of the algorithm are shown in Table I.

TABLE I. SIMULATION PARAMETERS OF LOGISTICS DATA WITH STATISTICAL DENSITY

Name	Value
P	50
Nc max	500
α	1
β	5
ρ	0.5

1) In terms of statistical concentration, other methods are used to generate a certain amount of pheromone at the starting point, and then the statistical density method is used to calculate, which improves the computational efficiency and accuracy.

2) The particle swarm optimization (PSO) algorithm is introduced to give each iteration a statistical concentration. Then, the solution generated by the statistical density system of each generation is used as the initial solution, and then the solution of other algorithms is repeated to speed up the calculation speed and achieve better results.

3) In the aspect of statistical density, the traditional selection method is based on experience, but if the selection is improper, it will reduce the operation speed, so other algorithms can be used to conduct.

4) The algorithm has the shortcomings of a premature algorithm, cannot adapt to the global optimal, is time-consuming, and so on. The statistical density algorithm is used to optimize the initial value and combined with other algorithms, such as crossover and mutation operation, to generate the optimal population.

B. Influence of Statistical Density on Data Mining of Logistics Frequent Paths

In the fundamental statistical density analysis, a pheromone is a positive feedback, and its change degree directly affects the degree of optimization of the research object. Therefore, this paper proposes a correction method based on positive and negative feedback to avoid excessive accumulation of pheromones.

In the research object's second iteration, its average path is calculated, and the updated pheromone standard formula is modified to make it move in a better direction, kf_{arg} . The statistical density is expressed by the average of the paths, as shown in the following Eq (11).

$$f_{avg} = \frac{1}{m} \bullet \sum_{k=1}^m L_i^k \quad (11)$$

The calculation formula in the following formula expresses the pheromone dynamic adjustment operator. Its

function is to adjust the pheromone concentration on the path in each iteration, Eq. (12).

$$\Delta_v = \sum_{i=1}^m \frac{L_i^k - f_{wrg}}{L_{best}^k} \bullet L_i^k \quad (12)$$

L_{best}^k It represents the shortest path and the path length passed by the current particle. The pheromones update Eq. (13) after adding the regulation operator is as follows: L_i^k

$$\tau_v(t+1) = (1-\rho)\tau_{ij}(t) + \Delta\tau_{ij} + \Delta_v \quad (13)$$

Optimising the statistical concentration in the iteration is easy because its calculation method is highly dependent on the amount of information. The particle swarm optimization algorithm calculation method uses the maximum value of Best and Pest to realize the particle location. The concept of Best and Pest is applied to the statistical concentration so that the research object has the characteristics of particles in the iteration, not only through the pheromone transmission but also through the optimization of the individual and the overall optimization. In the statistical density, the particles' adaptive crossover and mutation strategy prevents the method from falling into the maximum value. Firstly, the primary statistical concentration was used in the initial optimization. After passing through each channel, the FAG of the average road surface was calculated by using the statistical concentration, and the particles with fitness values below F were selected for cross-operation. When the fitness of path selection is not high, the position of the initial particle is replaced by the intersection point. Otherwise, it cannot be replaced. The crossover method randomly selects two locations, R and R, in a city and then makes a second crossover. Using local map technology, the contradiction between numbers can be effectively handled. Based on this, two arbitrary particles are exchanged to evaluate their adaptability, and when the particles' fit reaches an appropriate level, they will be replaced.

C. Effect of Statistical Density Algorithm on Data Simulation of Logistics Frequent Paths

Because the statistical density in this chapter is based on the minimization of the frequent person problem for logistics routing, it contradicts the up-and-down stochastic optimization principle of the benchmark function. The purpose of this paper is to evaluate the performance of the fusion algorithm in a specific city. The comparison method uses the last improved ACO and PSO and conducts ten separate experiments on different data to reduce the interference of random factors due to instability and record the best and average results. Table II illustrates the three methods for calculating the mean and the accuracy of the calculation, which are an example of computational density simulation.

For the "one-Kilometer" logistics problem, describe it as a multiple transportation point, that is, M mailers, N delivery points; this delivery point can be a user's address, can also be a smart ATM or a Courier, in this mode, all the information is considered as a job.

TABLE II. OPERATION TIME OF LOGISTICS DATA ALGORITHM BASED ON STATISTICAL DENSITY

Dataset	ACO	PSO	PSO-IACO
chn 31	22	17	10
ei1101	113	79	43
ch 130	253	189	78
rand 200	396	248	111

In the real situation, the "one-meter" distribution mode, under the premise of ensuring high quality and ensuring the delivery needs of customers, as far as possible to reduce the company's operating expenses. Therefore, applying the composite method to solve the terminal logistics problem still needs to conduct the division of the target area and route planning. Due to the current logistics terminal logistics environment in China, there are many kinds of couriers, and the target geographical distribution is wide; therefore, in the actual logistics process, it is inevitable to produce some problems such as going the wrong way, taking a long way, and repeating the road. If we cannot give an effective solution to this problem, there will be a lot of problems, such as unbalanced work, low efficiency, and repeated delivery in the end distribution process. Therefore, in the transportation of the "final kilometer", the "minimum route" is the primary optimization purpose, and then other restrictions are changed or added according to the actual situation to maximize the interests of customers and the company.

D. Optimizing Frequent Logistics Routes based on Statistical Density

Due to the use of multi-traveller mode, there will be energy limitations, capacity limitations, time limitations, and other issues. Energy limitation refers to the consumption of logistics and the continuous life of other intelligent vehicles. In the logistics process, the capacity limitation is the core of the vehicle loading capacity. In the case of delayed or early check-in by customers, there are also certain requirements on the delivery time of goods, namely the arrival sequence of the target. The "one kilometer of the city" problem to be dealt with in this paper should be modified or explained according to the actual situation to make the following interpretation:

1) *Description of the distance on the actual case:* The point of specific distribution and transport path between the plan because the curve of the road makes the actual transport distance cannot simply use coordinates to determine, in order to ensure the accuracy of the test and practical value, the point - point line spacing in the coordinate system to replace the actual distance of the target.

2) *Description of delivery personnel:* The delivery company is equipped with M delivery personnel, which can be assigned according to the task's scale.

3) *Description of time constraints:* Time constraints are added to each route. When the average speed of each train is the same, there is a positive relationship between the time and the total transportation distance. Therefore, in the future modelling process, in order to ensure real-time and balance, the maximum length of each route is limited.

IV. CONCLUSION

To sum up, this paper carries out three approaches for a single logistics problem: One is for a single logistics problem, using an improved statistical density and density method is optimized to it, and then USES the method of target classification, and then adopted based on the improved pheromone correction rule of statistical strength on the cluster planning, and the connection between these clusters established a whole calculation concentration simulation spacer ring. The transportation mode of the terminal is optimized. On this basis, the interval mode based on probability simulation cannot meet the transportation demand of the final kilometer, and the interval mode based on data density simulation can better meet the transportation demand of the terminal. Then, according to the "end" transportation situation in the logistics system, a M data distribution simulation region is established. In addition, the optimal combination of two different methods for distribution, on the premise of meeting the actual needs of customers, increasing the time limit, and making sure that the number of Courier service equalization effectively overcomes the traditional way of distribution, is not reasonable, the workload imbalance problem put forward an effective method of path selection, It has a practical guiding role in improving the customer's intimacy and improving the utilization rate of logistics.

REFERENCES

- [1] R. Wang, X. Li, Z. Zhang and M. Hongguang, "Modeling and simulation methods of sea clutter based on measured data", *International Journal of Modeling, Simulation, and Scientific Computing*, vol. 12, No. 1, 2050068 (2021).
- [2] J. Kearns, A. M. Ross, D. R. Walsh, *et al.*, "A blood biomarker and clinical correlation cohort study protocol to diagnose sports-related concussion and monitor recovery in elite rugby [Study Protocol]", *BMJ Open Sport & Exercise Medicine*, vol. 6, no. 1, 2020.
- [3] X. Yang, R. Zhang, F. Pan, *et al.*, "Stochastic user equilibrium path planning for crowd evacuation at subway station based on social force model," *Physical A: Statistical Mechanics and its Applications*, vol. 594, no. 127033, 2022.
- [4] B. Zhu, L. Zhang and Y. Zhang, "The Early Warning Method of Drug Adverse Reaction Monitoring Based on Data Mining Algorithm Was Studied," *Journal of Physics: Conference Series*, vol. 1852, no. 3, 2021.
- [5] F. Wu, M. Liu, G. Huang, *et al.*, "Simulation of stationary non-Gaussian multivariate wind pressures based on moment-based piecewise Johnson transformation model", *Probabilistic Engineering Mechanics*, vol. 68, no. 103225, 2022.
- [6] M. J. Marijnissen, C. Graczykowski and J. Rojek, "Simulation of the comminution process in a high-speed rotor mill based on the feed's macroscopic material data," *Minerals Engineering*, vol. 163, no. 106746, 2021.
- [7] T. Wisniewski and R. Szymański, "Simulation-based optimization of replenishment policy in supply chains," *International Journal of Logistics Systems and Management*, vol. 38, 2021.
- [8] Y. Zhao, H. Gao, S. Zhang, *et al.*, "Simulation and selection of fin stabilizers for polar cruise ships based on Computational Fluid Dynamics", *Journal of Physics: Conference Series* vol. 1, pp. 9, 2021.
- [9] C. Jin, J. Xiao and C. Cao, "Concept of biomimetic mechanical foot based on muscle simulation", *Journal of Physics Conference Series*, 1885(4):042017, 2021.
- [10] K. Ni and Y. Ma, "Simulation of LPG tank truck leakage and explosion accident based on CFD", *Journal of Physics: Conference Series*, vol. 2003, no. 1, 2021.

- [11] X. Wang and Y. Zhang, "A Data Simulation Method of Bank Fraud Transaction Based on Flow-Based Generative Model", *Journal of Physics: Conference Series*, vol. 1631, no. 1, pp. 1-9, 2020.
- [12] Q. Cheng, H. Shen, H. Chu, *et al.* "Research on Logistics Simulation and Optimization of Die Forging Production Line Based on Flexsim", *Journal of Physics Conference Series*, vol. 1624, no. 022063, 2020.
- [13] Z. H. Wang and T. Y. Li, "Computer optimal path simulation planning of oil and gas storage and transportation based on similarity measurement of decision space", *Journal of Physics: Conference Series*, vol. 1533, no. 3, pp. 5, 2020.
- [14] Y. Dong, L. Shao and J. Shi, "Study on Numerical Simulation of Roadway Backfill under Freeway Based on FLAC3D", *Journal of Physics: Conference Series*, vol. 1549, no. 4, pp. 6, 2020.
- [15] C. Yu, S. Cen and S. Luo, "Simulation design and optimization of production line of a cross-axis machining based on Plant Simulation", *Journal of Physics: Conference Series*, vol. 1654, no. 1, pp. 6, 2020.

Simulation Analysis of Hydraulic Control System of Engineering Robot Arm Based on ADAMS

Haiqing Wu

Department of Mechanical Engineering, Taiyuan Institute of Technology, Taiyuan, Shanxi, 030008, China

Abstract—Substantial trenching capacity, communication capabilities, simple configuration, and so on are just a few of the many benefits that make Hydraulic Control Systems (HCS) the context of physical devices used within the geotechnical trench. These characteristics have led to widespread application in developing water conservation and hydroelectric technology, architectural construction, local construction, and other technology. In this article, the engineering robot arm proposed an HCS. Subsequently, a digital version of the functional device is constructed using Anti-Doping Administration and Management System (ADAMS), a simulation program, by incorporating associated restrictions and workload. With the help of a simulation model of the HCS's functioning apparatus, this research obtains the fundamental factors of the excavator's operating range and the pressured condition variation curve of the location of every Hydraulic Actuator (HA). The findings, which provide a conceptual framework and enhancements for the control system equipment, significantly raise the bar on China's excavator architecture, expand digger efficiency, and foster the firm's fast growth. An in-depth examination of the HCS's current operating condition, including an examination of the simulated model's transmission phase, can be determined. The findings provide a theoretical foundation for designing an optimal HCS.

Keywords—Hydraulic control systems; ADAMS; simulation analysis; engineering robot arm

I. INTRODUCTION

Being a crucial component of automated control technology, digital electric Hydraulic Control Systems (HCS) have found several uses in numerous industries, including aviation, power generation, and more. An essential component of any digital HCS, electrical hydraulic gates convert low-voltage electric impulses into elevated hydraulic energy, making them essential for any application requiring electrical hydraulic servo control. So, if the electro-hydraulic control circuit valve had failed, the dependability of systems like the electromagnetic flow HCS and the automated control system would have been affected. That's why it's important to focus more on things like doing a technical failure study on the electromagnetic, hydraulic servo valves to make sure the electro-HCS is reliable and will last for the long haul [1]. Demolition robots are cutting-edge tools for the modern deconstruction of reinforced concrete structures. It has several applications in fields as diverse as nuclear energy, disaster response, metalworking, and demolition. The booms arm, which comprises three arm parts and then a crushing hammer mechanism, is the most important part of a destructive robot. Each device contains a hydraulic cylinder connected to a mechanical arm or breaker hammer, depending on the

application. The reliability and safety of the building, as well as the profitability and expense of the enterprise, are all impacted by the vibrational properties. Thus, the arm's construction is crucial. As a result, the reliability of components like the automated management system and the digital electric HCS might have been compromised if the electro-hydraulic circuitry valve had failed [2]. Individuals with amyotrophic lateral sclerosis, spinal cord injuries, and other illnesses that cause paralysis from the neck down sometimes wholly or partly lose control of their limbs. Certain activities of daily living, such as getting from one location to another or grasping a glass to drink from, are impossible for paralyzed individuals to do without external support. Providing such aid often requires a nurse or other paramedic, which might take a long time and need many resources. Artificial technologies, such as robotic arms, are often exploited to enable two fundamental and necessary limb tasks for paralyzed people to care for themselves: moving and gripping. Paralyzed people may do everyday chores with a robotic arm that mimics arm motions and restores the patient's ability to hold objects.

In this, a robotic arm was used to pour coffee from a bottle using the patient's motion intents [3]. The primary focus of these control systems is on the robotic arm's redundancy resolution. Plans for the robotic arms and non-holonomic vehicle movements are provided. Priorities of these intersecting paths are described in more detail in the article. If the automobile or the robotic arm has trouble following the desired trajectory, the control algorithm will prioritize which system will complete the job. In this, the authors simulate and execute the dual control of a robotic arm placed on a wheelchair. The redundant arm's calculated configuration is the outcome of an optimization exercise. Hydraulic power robotic manipulators are analyzed and controlled in a separate field [4]. Fig. 1 depicts the Hydraulic Robot System (HRS) block diagram.

The robotic arm's redundant resolution is the main emphasis of these control systems. Compared to electric motors, they still offer superior power-to-weight ratios and more inherent toughness and rigidity. Academics are paying more attention to creating high-performance controllers for applications such as footed robot control, actuator impedance control, and flight simulator motion control, where precise control efficiency is essential. The data is interpreted by Hydraulic Actuators (HA) as a velocity instruction rather than a force instruction like their electric equivalents, complicating the difficulty of HRS. As a result, the well-researched generic industrial robot approaches cannot be directly used [5].

Typically, robot arms are computationally designed to carry out instructions. Devices like delivery ordering systems, robot arms, and automated guided vehicles may all benefit from being part of a unified network and being directed by AI in an entirely autonomous factory. Some examples of these advantages include higher production and productivity and a better investment return for gear. Microsoft, Amazon, and Google are just a few companies that provide developers with the frameworks, tools, and platforms necessary to train AI [6].

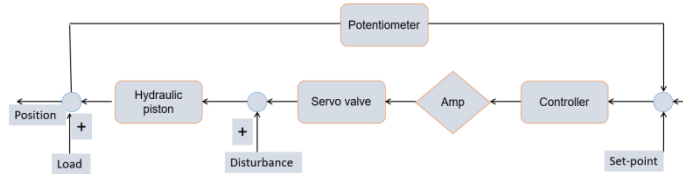


Fig. 1. Block diagram of the HRS.

Compared to contemporary industrial counterparts, non-industrial robotic arms are flexible, have little actuation power, and may include components that display huge variances concerning respective specifications, all of which achieve extensive trajectory tracking and significant difficulty. It restricts the use of flexible or low-cost robots to activities that need great precision or places where circumstances may rapidly change, such as on building sites, where robots are subjected to extreme weather and working circumstances. Operating robots at low speeds, where the dynamics coupling among joints is minimal and is a straightforward solution to cope with such situations, but it comes at the cost of reduced performance and output. Using the maximum potential of cooperative and low-cost robots outlines strategies for improving accuracy and accelerating the process in this objective [7]. Markets for robotics technology are expected to expand rapidly over the next decade, and the introduction of robotics will cause a revolution in the heavy machinery sector, just as it has already done in other areas, such as transportation and manufacturing automobiles. To this day, energy efficiency is still one of the most pressing problems to be addressed in hydraulic systems. The preceding intellectual HRS are inefficient since they use a standard, ineffective valve control. Energy efficiency may be less of a priority during the design phase for fixed installations.

On the other hand, ambulatory robotic systems have a unique challenge because they have very little storage room for their energy sources [8]. Motion sensors attached to the machinery and the human body allow the arm motions and prosthesis to be controlled by inbuilt ultrasonic sensors, somatosensory rhythms, and attempting movements. This approach, however, may lead to significant control errors in the grip and requires frequent changes to the sensor values. As neuromuscular stimulating and intrusive BCIs have advanced in the biomedical field, robotic arms have been programmed to reach out and grab objects. Surface electromyographic and non-invasive EEG inputs have been successfully used for robotic arm gripping [9]. Existing arm-grabbing systems may be classified into non-computer vision-based and machine-learning systems. Robotic arms are often used for grasping. Ultrasonic sensors, central nervous system connections, and electromyograms are only a few of the non-computer vision-

based devices that benefit arm gripping. Moving sensors attached to the computer and the human condition allow the arm actions and prosthesis to be controlled by internal ultrasonic sensors and attempting movements.

II. RELATED WORKS

The best material for the manufacturing robotic arm has to be chosen to achieve the necessary qualities at the lowest possible cost and with the most potential for the intended use. The paper uses the analytical hierarchical process technique to decide which materials to apply in a robotic system [10]. One of the most popular multiple-criteria decision approaches, the Analytic Hierarchy Process (AHP), considers multiple criteria simultaneously. The AHP method's strength lies in its ability to accommodate numerous criteria simultaneously [11]. To solve the challenge of optimizing the design of robot arms for high-speed performance, we offer a surrogate-based evolutionary optimization technique using a global optimization technique, combining the response surface technique with a multi-objective evolutionary computation via decomposition. First, the robot arm's architecture and performance are evaluated using CAD software like Inventor and finite element modelling software like ANSYS [12]. This paper introduces a novel approach to teaching robotics using four components: an algorithm, a virtual experiment, some programming, and a controller.

The authors present an inverse kinematics method with an inexpensive desktop six-axis robotic arm designed for educational purposes. In MATLAB, we implement the inverse solution technique, and in VREP, we create a dynamic model of the desktop robot's arm (V-REP). By moving the virtual robot arm in V-REP to the desired location and orientation using MATLAB's API interface, we can test the efficacy of the robot manipulator technique and ensure its accuracy [13]. The research examines two degrees of freedom (DOF) robot navigation and builds free Cartesian spaces for finding the space available that ensures a collision-free route. An improved Ant Colony Optimization (ACO) method is offered to achieve motion goal-congruent optimum route planning. The improved ACO algorithm's primary purpose is to provide an optimum route based on the precise distance of the D* method. Integrating polynomial equations of the fifth order produces a path with minimal transitional points between the starting and ending points [14]. To improve the trajectory tracking control, the study presents two strategies for dealing with singularities that may arise when a robot arm follows a particular path. The paper uses genetic algorithms, but one is more localized and the other more worldwide. Each approach was tested on polynomial trajectories up to the third degree, and their results were compared based on metrics like trajectory inaccuracy, number of singularities, and computational cost. According to the outcomes, the technique based on the global genetic algorithm performed the best, as it could trace the second and third-degree trajectory with the fewest errors, singularities, and computational costs [15].

They suggested a Deep Learning (DL) model based on multi-directional Convolutional Neural Networks (CNN) with bidirectional Long Short-Term Memory (LSTM). The autocorrelation and the normalization root mean squaring

error were used to evaluate the decode performance for different directions in 3D space [16]. The hardware includes the robot's design process, motor selection, and electrical components used to control the robot's joints. The software comprises algorithms that manage the robot's movements to ensure it moves according to specifications and algorithms that transform needed words into proper sequences with target areas. In this scenario, voice recognition software serves as a guide throughout the writing process [17]. Through the use of a small sample of productive results and subsequent actively tendered queries, in which the robot displays a state and starts asking for a label to evaluate whether or not that process and system accomplished the assignment, they propose an approach to eliminate the need for manual designing of reward specific requirements. Instead of expecting the user to personally give the reward signal by labelling every condition encountered during training, our technique only needs labels for a small subset of states, providing a feasible and effective method for learning new abilities without creating incentives [18] manually.

The strategy relies on a sampling approach and has two main components. When a starting point has been found by web search, a greedy approach is used to optimize the route by locally applying adaptive filters to the parts of the path that have particularly severe jerks. Numerical optimization is used to produce the filtered outcome. Developing a collision-indication function expressed as a support-vector machine is more computationally efficient using an adaptable sampling strategy [19]. The study evaluates AL against the most popular data sampling techniques for predicting regression outcomes with reduced sample sizes. The paper provides a unique assessment framework for comparing alternative sampling strategies in a regulated and objective way, despite their varied needs. They examine the sampling efficiency, stability, and predictive value of the resultant ML models from three illustrative use cases (UCs) to determine whether AL or DOE approaches are preferable for data production [20].

III. MATERIALS AND METHODS

A. ADAMS Modeling and Simulation Setup

Using a particular professional environment that may replicate the actual environment down toward the units, gravitation, and operating grids, the HCS design could be loaded into ADAMS. One of the first things to finish after a transfer should be to add restrictions and motor features, including a fixed pair between the land surface and the core, a spinning pair, a block pair, and a roller bearings joint pair at every section of the connection, and a mobility pair between the HA and it is relating engine shaft. Next, users must recognize the step response as the driving factor behind the engineering robot arm. Finally, execute the simulation via the modelling and analysis tool after all the preceding steps. Make sure there are no extraneous restrictions in the system and that there are no free variables.

B. Parameters of a Functional Hydraulic Cylinders Motion

The lifting arm hydraulic cylinder length L1, the buckets pole hydraulic cylinder lengths L2, and the bucket hydraulic cylinder length L3 all play a role in establishing the precise

geometries of a spade. After L1, L2, and L3 are set in stone, it's clear where the anti-shovel mechanism will be placed. Anti-shovel hydraulic cylinder motion characteristics are shown in Table I.

C. Movement Function and Trajectory Simulations of a Hydraulic Cylinder

The following is a motion simulator of a crane hydraulic cylinder, a bucket rod hydraulic cylinder, and a bucket hydraulic press.

1) *Hydraulic cylinder boom motion model:* Full-shrinkage adjustment cylinder bucket rod position. Make sure the cylinder is set to full contraction before adjusting the bucket. The hydraulic cylinder's motion function allows the boom to extend and retract. To play back simulation results and create curves, you must transfer ADAMS/Post-processed Modular. Trace Marker could trace the path of the excavator's bucket's tooth marking points, and the resulting trajectory chart is shown in Fig. 2.

2) *Hydraulic cylinder and bucket rod model in motion:* If the bucket tooth is a direct line, then the Moving function entails changing the connection between the boom and the bucket rod and the connection between the bucket rod and the bucket. To play back simulation results and create bends, it must convert ADAMS/Post-processed components.

3) *Hydraulic cylinder movement simulator for a bucket:* The excavation work trajectory of a resource extraction bulldozer is a circular arc, with the steak between the bucket and rod the bucket serving as the middle of the plot line and the proximity from the tendon to the bucket tooth serving as the radius; the wrap edge the arc length are both calculated by the stroke of the bucket hydraulic cylinder.



Fig. 2. An animated model of manipulating arms and a model of a bucket rod.

D. Dynamic Simulation Analysis

Regarding excavators, the primary area of investigation is the force fluctuation at the bucket's teeth and hinge point, which may be gleaned through a simulation model of the machine's mechanical power relations when subjected to

different external loads. The goal is to guarantee that the design will work as intended.

1) *Estimating Work in a Complicated Mining Procedure:* Friction and soil resistance is not considered. Throughout its time spent underground, an excavator is subjected to those mentioned above primary external loads:

a) *Mines that provide normal and tangential resistance:* Mining resistance may be considered operating just on top of the bucket tooth in two directions: the tangential direction, which is perpendicular to the tracks, and the regular direction, which is perpendicular to the track. The following is their EQU (1) based on experience:

$$\begin{aligned} W_1 &= K_0bh \\ W_2 &= \psi W_1 \end{aligned} \quad (1)$$

The resisting ratio of k_0 mining is representative of the overall resistance encountered by bucket-style excavating tools underground. Failure of the soil, soil loading, and friction all contribute to the overall resistance. The SI unit for k_0 N/cm^2 Miniature excavators often functions best on flat ground. Based on a table, we get a $k_0=15N/cm^2$ b-cutting width value as unit cm. Based on this model, $b=60$ cm. The unit for h-cutting depth is centimetres. On average, $h=0.2$ $b=12$ cm. ψ resistant to Mining Parameters. Accessing a lookup table reveals, $\psi = 0.6$. The following information is obtained by plugging the figures as mentioned above into EQU (2), and computing

$$\begin{aligned} W_1 &= 10800, N = 10.8KN \\ W_2 &= 0.5, W_1 = 5.4KN \end{aligned} \quad (2)$$

b) *Resistance to lifting:* After excavating is complete, hoisting resistance mainly pertains to the gravitation of the elements in the bucket. It is attributed to the bucket's centroid as normal and always moves vertically downwardly.

$$G = \rho V_g = rV \quad (3)$$

The capacity of a V-Bucket: $V = 0.28m^3$ in line with the value of the specification. Level dirt has a density of $1.8 \times 10^4 kg / m^3$. Based on the gravity setting in ADAMS, it is given by $g = 9.8m / s^2$. The gravitational value of the components may then be determined around $5KN$.

c) *Load up the vehicle and enter ADMAS:* The acute and normal directions of the bucket teeth match the tangential and normal directions of resistance. It demonstrates that these forces all act in the same direction. We can calculate the variations in the force of each cylinder at every hinge point whenever it works in the deep position by applying a typical complicated motion in excavation as a work circle. We may

also get the excavation's kinetic models without considering the response times and the length of the working cycle, which includes excavation and unloading. The pace of bucket digging often determines how long it takes to excavate. The running speed is set to 0.5 m/s, the digging speed to 0.75 m/s, and the unloading time to 2s to make the simulation easier.

We split a cycle into three parts and determined the overall cycle duration to be 8s according to the results above:

(i) *The bucket phase that descends:* We adapt the working device's digging point to the broadest possible digging diameter, which entails reducing the size of the bucket rod and boom cylinder. We presume no additional loads are being imposed throughout this operation.

(ii) *The phase of digging:* The bucket and bucket rod cooperate to fill the bucket with earth. It denotes the end of the digging process. The load in this operation is the normal and tangential digging resistance. They start off increasing as the bucket angle rises. Digging resistance increases the most while digging the deep and decreases when the bucket angle increases.

(iii) *Phases of hoisting and unloading:* During the hoisting phase, we first set the boom cylinder to its full shrinking position while adjusting the bucket cylinder to guarantee that items are not dispersed throughout the ascent in the bucket. Then finish emptying, and adjust the bucket rod and cylinder to the complete shrinkage condition. The primary burden in the mining process is the weight of the material, and the boom-raising stage does not change as the weight of the material increases.

IV. RESULT AND DISCUSSION

To provide the required HCS, the HCS first pumps the hydraulic fluid, which then circulates across the device. The hydraulic mechanism allows the bigger versions of the commercial or hydraulic robotic arm to carry substantial weights; hence, the arms are typically metal. Findings from the engineering robotic manipulator for the ADAMS model are analyzed below. Metrics such as accuracy, precision, recall, mobility, and energy transformation are compared with the conventional methods. The conventional methods are Artificial Intelligence (AI) [21] and DL [22]. The findings are listed below.

A. Accuracy

As a measure of a device's accuracy, it refers to how closely calculated values align with the real thing. It has been shown that the proposed approach is more precise than the existing method. We provide our results in terms of the percentage of accuracy. In Fig. 3, the recommended system's accuracy is shown. AI has achieved 75% accuracy, but ML has only reached 83%. The proposed method has 95% of accuracy. It demonstrates that the proposed method is superior to the traditional methods. Table I represents the accuracy.

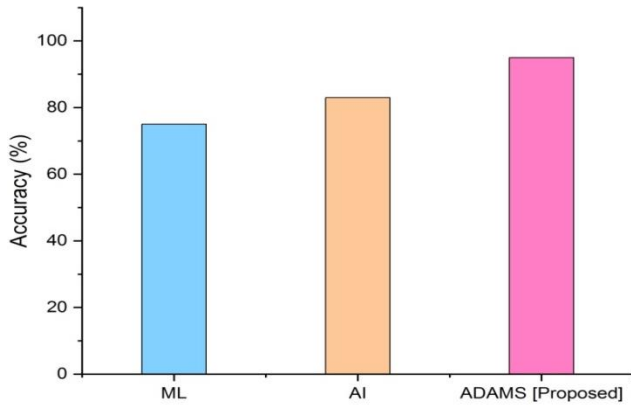


Fig. 3. Accuracy.

TABLE I. ACCURACY MEASUREMENT

Techniques	Accuracy (%)
ML	75
AI	83
Proposed ADAMS	95

B. Precision

Precision, often called positive prediction value, is the percentage of correct opinions among the retrieved instances. Precision is a determinant of value, which it may determine. Precision measures how likely it is that a specific recovery will occur. The proposed work has significantly higher accuracy than the current methods. Fig. 4 shows the results of comparing the precision of conventional and proposed methods. Conventional methods achieve the following levels of accuracy; therefore, ML has attained 77%, and AI has reached 83%. As a result, the proposed system offers the highest potential outcome of 95%. The precision value is shown in Table II.

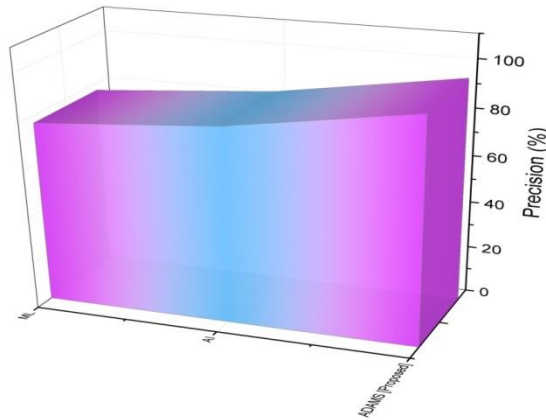


Fig. 4. Precision.

TABLE II. PRECISION VALUE

Techniques	Precision (%)
ML	77
AI	83
Proposed ADAMS	95

C. Recall

Fig. 5 compares the proposed technique and the existing method for recall. The recall is the fraction of relevant events which have been retrieved. One other name for sensitivity is recalled. The proposed method has the greatest recall of all the conventional systems. Conventional methods' recall for predicting performance is as follows: ML achieves 73%, AI achieves 85%, and the proposed method achieves 97%. It shows that the intended work will be completed effectively. The recall analysis is shown in Table III.

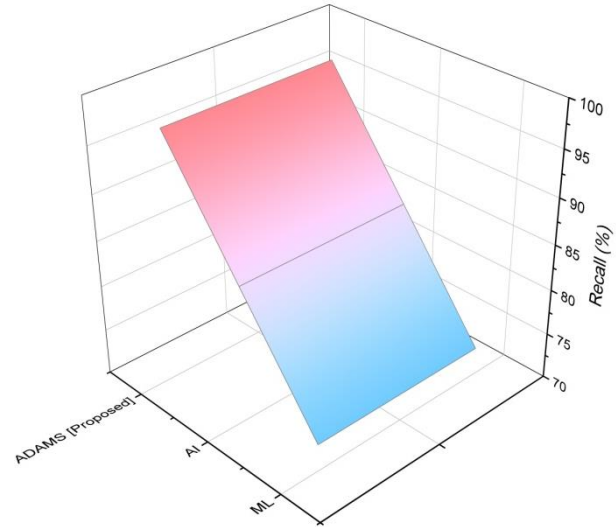


Fig. 5. Recall.

TABLE III. RECALL

Techniques	Recall (%)
ML	73
AI	85
Proposed ADAMS	97

D. Mobility

Robotics designed for long-distance transportation are called Mobility Systems. The ability to move freely is crucial for automated systems to carry out their missions in challenging and intricate settings. Wheels, feet, hops, and other forms of mobility may all be used by robotic systems. The mobility contrast is demonstrated in Fig. 6. Established methods like ML have attained 75% mobility, while AI has reached 81% mobility. According to the findings, the proposed system is mobility towards the extent of 93%. The results of the mobility research are presented in Table IV.

TABLE IV. MOBILITY

Techniques	Mobility (%)
ML	75
AI	81
Proposed ADAMS	93

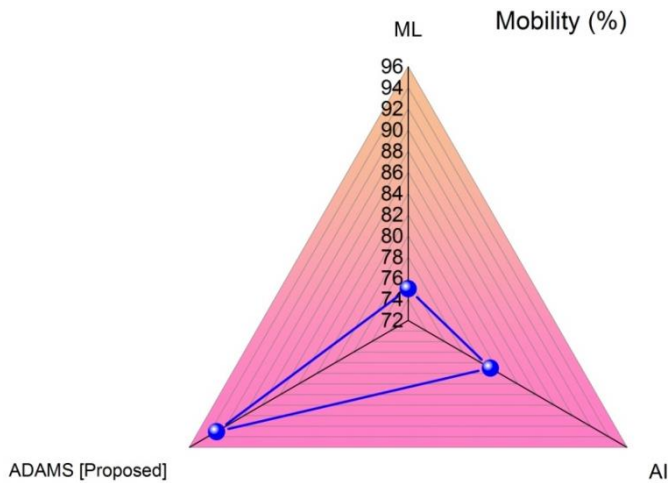


Fig. 6. Mobility.

E. Energy Transmission (ET)

Converting one type of resource into another is termed ET. The contrast between the two forms of ET is seen in Fig. 7. Existing methods, such as ML, have attained a mobility rate of 73%, and an AI of 88% in the ET. The results show that the proposed method can convert energy for robotics technology by a significant percentage is 97%. The analysis of ET is shown in Table V.

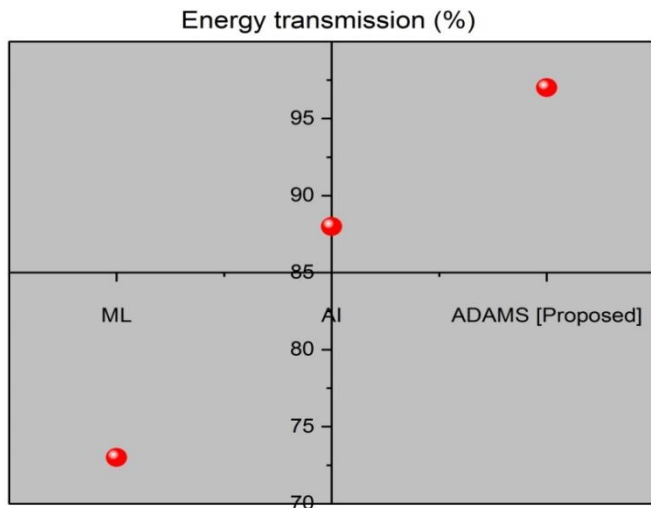


Fig. 7. Energy transmission.

TABLE V. ENERGY TRANSMISSION

Techniques	ET (%)
ML	73
AI	88
ADAMS [Proposed]	97

V. CONCLUSION

This study uses ADMAS software's robust analysis feature to examine the engineering robot's simulation Hydraulic Control System (HCS)'s functioning device. We have determined the scope of this research work and the significant

fundamental characteristics of accuracy, precision, recall, mobility, and Energy Transmission (ET), and we have acquired the defined points and the force curve of a Hydraulic Actuator (HA) via a simulation model. This study obtained a model simulation study of the HCS's operational state, and our findings may be used as a theoretical foundation for the development and optimization of engineering robots' HCS.

ACKNOWLEDGMENT

Joint Science and Technology Innovation Project of Colleges and Universities in Shanxi Province (2020L0642).

REFERENCES

- [1] T.W. Ni and Z.G. Yang, Failure analysis on unexpected leakage of electro-hydraulic servo valve in digital electric hydraulic control system of 300 MW thermal power plant”, Engineering Failure Analysis, vol. 119, pp.104992 2021.
- [2] J. Li, K. Wang, K. Zhang, Z. Wang and J. Lu, “Design and analysis of demolition robot arm based on finite element method”, Advances in Mechanical Engineering, vol. 11, no. 6, 2019.
- [3] Q. Huang, Y. Chen, Z. Zhang, S. He, R. Zhang, J. Liu, Y. Zhang, M. Shao and Y. Li, “An EOG-based wheelchair robotic arm system for assisting patients with severe spinal cord injuries”, Journal of neural engineering, vol. 16, no. 2, pp.026021, 2019.
- [4] B. Varga, S. Meier, S. Schwab and S. Hohmann, “Model predictive control and trajectory optimization of large vehicle manipulators”, IEEE International Conference on Mechatronics (ICM), vol. 1, pp. 60-66, 2019.
- [5] Y. Huang, D.M. Pool, O. Stroosma and Q. Chu 2019, “Long-stroke hydraulic robot motion control with incremental nonlinear dynamic inversion”, IEEE/ASME Transactions on Mechatronics, vol. 24, no. 1, pp.304-314, 2019.
- [6] M. Matulis and C. Harvey, “A robot arm digital twin utilizing reinforcement learning”, Computers & Graphics, vol. 95, pp.106-114, 2021.
- [7] A. Carron, E. Arcari, M. Wermelinger, L. Hewing, M. Hutter and M.N. Zeilinger, “Data-driven model predictive control for trajectory tracking with a robotic arm”, IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 3758-3765, 2019.
- [8] J. Koivumäki, W.H. Zhu and J. Mattila, “Energy-efficient and high-precision control of hydraulic robots”, Control Engineering Practice, vol. 85, pp.176-193, 2019.
- [9] X. Chen, B. Zhao, Y. Wang and X. Gao, “Combination of high-frequency SSVEP-based BCI and computer vision for controlling a robotic arm”, Journal of neural engineering, vol. 16, no. 2, pp.026012, 2019.
- [10] B. Cheng, W. Wu, D. Tao, S. Mei, T. Mao and J. Cheng, “Random cropping ensemble neural network for image classification in a robotic arm grasping system”, IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 9, pp.6795-6806, 2020.
- [11] A. Kumar and M. Kumar, “Implementation of analytic hierarchy process (AHP) as a decision-making tool for the selection of materials for the robot arm”, International Journal of Applied Engineering Research, vol. 14, no. 11, pp.2727-2733, 2019.
- [12] J.C. Hsiao, K. Shivam, C.L. Chou and T.Y. Kam, “Shape design optimization of a robot arm using a surrogate-based evolutionary approach”, Applied Sciences, vol. 10, no. 7, pp.2223, 2020.
- [13] D. Zhou, M. Xie, P. Xuan and R. Jia, “A teaching method for the theory and application of robot kinematics based on MATLAB and V-REP”, Computer Applications in Engineering Education, vol. 28, no. 2, pp.239-253, 2020
- [14] A.T. Sadiq, F.A. Raheem, and N. Abbas, “Ant colony algorithm improvement for robot arm path planning optimization based on D* strategy”, International Journal of Mechanical & Mechatronics Engineering, vol. 21, no. 1, pp .96-111, 2021.

- [15] P. P. Reboucas Filho, S. P. P. da Silva, V. N. Praxedes, J. Hemanth and V.H.C. de Albuquerque, "Control of singularity trajectory tracking for robotic manipulator by genetic algorithms", *Journal of computational science*, 30, pp.55-64, 2019.
- [16] J. H. Jeong, K. H. Shim, D. J. Kim and S. W. Lee, "Brain-controlled robotic arm system based on multi-directional CNN-BiLSTM network using EEG signals", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp.1226-1238, 2020.
- [17] S. Yuvaraj, A. Badholia, P. William, K. Vengatesan and R. Bibave, "Speech Recognition Based Robotic Arm Writing. In Proceedings of International Conference on Communication and Artificial Intelligence: ICCAI 2021", pp. 23-33, Singapore: Springer Nature Singapore, 2022.
- [18] A. Singh, L. Yang, K. Hartikainen, C. Finn and S. Levine, "End-to-end robotic reinforcement learning without reward engineering", arXiv preprint arXiv:1904.07854, 2019.
- [19] C. Dai, S. Lefebvre, K.M. Yu, J.M. Geraedts and C.C.Wang, "Planning jerk-optimized trajectory with discrete time constraints for redundant robots", *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 1711-1724, 2020.
- [20] M. Amini, K. Sharifani and A. Rahmani, "Machine Learning Model Towards Evaluating Data gathering methods in Manufacturing and Mechanical Engineering", *International Journal of Applied Science and Engineering Research*, vol. 15, pp. 349-362, 2023.
- [21] J. Wu, H. Shen, T. Liu, J. Cai, F. Duan and P. Dong, "Simulation Analysis of Hydraulic Synchronous Upside Down Hoisting Crossing Device for Transmission Line Based on Artificial Intelligence", *IEEE International Conference on Knowledge Engineering and Communication Systems (ICKES)*, pp. 1-5, 2022.
- [22] K. Huang, S. Wu, F. Li, C. Yang and W. Gui, "Fault diagnosis of hydraulic systems based on deep learning model with multi-rate data samples", *IEEE Transactions on neural networks and learning systems*, vol. 33, no. 11, pp.6789-6801, 2021.

Enhanced Transfer Learning Strategies for Effective Kidney Tumor Classification with CT Imaging

Muneer Majid^{1*}, Yonis Gulzar^{2*}, Shah Nawaz Ayoub³, Farhana Khan⁴,
Faheem Ahmad Reegu⁵, Mohammad Shuaib Mir⁶, Wassim Jaziri⁷, Arjumand Bano Soomro⁸
Glocal School of Science and Technology, Glocal University, Delhi-Yamunotri Marg (State Highway 57),
Mirzapur Pole, Dist - Saharanpur, U.P. - 247121, India^{1,3,4}
Department of Management Information Systems, College of Business Administration,
King Faisal University, Al-Ahsa 31982, Saudi Arabia^{2,6,7,8}
College of Computer Science and Information Technology, Jazan University, Jazan 45142, Saudi Arabia⁵
Department of Software Engineering-Faculty of Engineering and Technology, University of Sindh, Sindh, Pakistan⁸

Abstract—Kidney tumours (KTs) rank seventh in global tumour prevalence among both males and females, posing a significant health challenge worldwide. Early detection of KT plays a crucial role in reducing mortality rates, mitigating side effects, and effectively treating the tumor. In this context, computer-assisted diagnosis (CAD) offers promising benefits, such as improved test accuracy, cost reduction, and timesaving compared to manual detection, which is known to be laborious and time-consuming. This research investigates the feasibility of employing machine learning (ML) and Fine-tuned Transfer Learning (TL) to improve KT detection. CT images of individuals with and without kidney tumors were utilized to train the models. The study explores three different image dimensions: 32x32, 64x64, and 128x128 pixels, employing the Grey Level Co-occurrence Matrix (GLCM) for feature engineering. The GLCM uses pixel pairs' distance (d) and angle (θ) to calculate their occurrence in the image. Various detection approaches, including Random Forest (RF), Support Vector Machine (SVM), Gradient Boosting (GB), and Light Gradient Boosting Model (LGBM), were applied to identify KT in CT images for diagnostic purposes. Additionally, the study experimented with fine-tuned ResNet-101 and DenseNet-121 models for more effective computer-assisted diagnosis of KT. Evaluation of the efficient diagnostics of fine-tuned ResNet-101 and DenseNet-121 was conducted by comparing their performance with four ML models (RF, SVM, LGBM, and GB). Notably, ResNet-101 and DenseNet-121 achieved the highest accuracy of 94.09%, precision of 95.10%, recall of 93.5%, and F1-score of 93.95% when using 32x32 input images. These results outperformed other models and even surpassed state-of-the-art methods. This research demonstrates the potential of accurately and efficiently classifying KT in CT kidney scans using ML approaches. The use of fine-tuned ResNet-101 and DenseNet-121 shows promising results and opens up avenues for enhanced computer-assisted diagnosis of kidney tumors.

Keywords—Kidney; kidney tumor; automatic diagnosis; machine learning algorithms; CT imaging; deep learning; transfer learning

I. INTRODUCTION

The kidneys, two bean-shaped organs located on either side of the spine, are indispensable for maintaining the body's internal equilibrium and overall health. These vital organs play a pivotal role in filtering waste products, excess salts, and

water from the blood, while also regulating blood pressure and producing essential hormones for red blood cell production. With such critical functions, any disruption or abnormality in kidney health can have serious repercussions on an individual's well-being [1]. Kidney tumors can be benign or malignant [2]. The benign kidney tumor can be a cyst, masses, or lipoma whereas malignant kidney tumor refers to renal cancer, pelvic cancer [3]. Kidney cancer, characterized by the presence of tumors arising from renal tissues, poses a significant global health challenge. Despite extensive research, the precise etiological factors triggering kidney cancer remain enigmatic, though hereditary and environmental influences are among the potential contributors. The insidious nature of kidney tumors often leads to asymptomatic progression, delaying diagnosis until advanced stages.

In 2023, it is estimated that 81,800 adults in the United States will be diagnosed with kidney cancer. Kidney cancer is more common in men, and the average age of diagnosis is 64, with most cases occurring between ages 65 and 74. The number of new kidney cancer cases has been increasing, though the rate of increase has slowed in recent years, partially due to increased use of imaging tests that can detect small kidney tumors incidentally [4]. In 2020, an estimated 179,368 people worldwide died from kidney cancer. The 5-year relative survival rate for kidney cancer in the United States is 77%. This rate varies depending on cancer stage, age, general health, and treatment effectiveness. For instance, the five-year relative survival rate is 93% for those with cancer confined to the kidney, 72% if cancer has spread to surrounding tissues or lymph nodes, and 15% if it has spread to distant parts of the body [4].

In the realm of medical imaging, Computed Tomography (CT) scans, ultrasounds, and Magnetic Resonance Imaging (MRI) serve as vital techniques, affording physicians a comprehensive visualization of the kidneys and associated tumors, thereby enabling meticulous evaluation of their dimensions, morphology, and characteristics [5,6]. Despite their undeniable utility, distinguishing between healthy tissue and malignant growth in kidney scans poses a formidable challenge. While manual diagnosis and expert detection of kidney tumors boast commendable accuracy, they demand significant time and effort [7], and results may exhibit

*Corresponding Author: ygulzar@kfu.edu.sa, muneerbazaz@gmail.com

variability amongst different practitioners. The paramount significance of early detection and accurate classification of tumors lies in mitigating the risk of metastasis to other anatomical regions. As an expedient and efficient alternative, automatic detection emerges as a promising approach, streamlining diagnostic processes and potentially safeguarding patients' lives. Although manual detection is renowned for its precision, the automatic diagnostic methodology offers expedited results, without compromising comparability to manual findings.

Artificial intelligence (AI) and deep learning have revolutionized various industries, including agriculture [8–13], education [14, 15], finance [16], and healthcare [17–19]. In the field of healthcare, AI has shown tremendous promise in improving patient outcomes, enhancing diagnostics, and streamlining healthcare processes. With the ability to analyze vast amounts of data and identify complex patterns, AI-powered systems have opened new frontiers for early disease detection, personalized treatment plans, and overall healthcare efficiency. In healthcare, one of the areas where AI and deep learning have made significant advancements is in the early detection of diseases, including cancer [20]. Detecting cancer at an early stage is crucial for improving treatment success and patient survival rates. Kidney cancer, for example, often presents with few symptoms in its early stages, making early detection challenging. However, deep learning algorithms have proven to be effective in analyzing medical imaging data, such as CT scans and MRI images, to detect kidney tumors at their nascent stages [21].

From the literature, it is evident that many researchers have incorporated deep learning in classifying kidney tumors. Lee et al. [22] developed an automatic deep feature classification (DFC) method using hand-crafted and deep features, along with machine learning classifiers, to distinguish benign angiomyolipoma without visible fat (AMLwvf) from malignant clear cell renal cell carcinoma (ccRCC) in abdominal contrast-enhanced computer tomography (CE CT) images. The proposed method achieved an accuracy of $76.6 \pm 1.4\%$ using the combination of hand-crafted and deep features, outperforming HCF-only and DF-only methods by 6.6%p and 8.3%p, respectively. Texture image patches (TIPs) were introduced to emphasize texture information and reduce mass size variability, resulting in steady performance regardless of the convolutional neural network (CNN) models used. Han et al. [23] used an image-based deep learning framework to classify renal cell carcinoma subtypes using CT images. The neural network achieved 0.85 accuracy, 0.64-0.98 sensitivity, 0.83-0.93 specificity, and 0.9 AUC, showing promising results for subtype classification and potential clinical cooperation with radiologists. Deep learning framework achieved 93.39% accuracy in classifying clear cell RCC and 87.34% for chromophobe RCC from histopathological images. A novel support vector machine-based method improved classification accuracy to 94.07% for distinguishing clear cell, chromophobe, and papillary RCC. The CNN also extracted morphological features to predict patient survival outcome, showing potential for cancer diagnosis and prognosis [24]. Oberai et al. [25] developed a semi-automated majority voting

CNN-based method to classify renal cell carcinoma (RCC) from benign solid renal masses on contrast-enhanced computed tomography (CECT) images. The CNN model achieved 83.75% accuracy in differentiating RCC from benign masses. A fully automated approach yielded 77.36% accuracy, while a 3D CNN achieved 79.24% accuracy in renal mass classification. Pedersen et al. [26] used a modified version of ResNet50V2 CNN to differentiate oncocytoma from renal cell carcinoma (RCC) using non-invasive imaging. They collected 20,000 2D CT images from 369 patients for training, validation, and testing. The model achieved 93.3% accuracy and 93.5% specificity on the main test set and 90.0% accuracy and 98.0% specificity on the additional validation set. When evaluated with a majority vote for each patient, the accuracy rose to 100%, reducing false negatives to zero, demonstrating the potential of CNNs for accurate diagnosis.

Sudharson et al. [27] proposes an automatic classification method for B-mode kidney ultrasound images using an ensemble of deep neural networks (DNNs) with transfer learning. The DNNs, including ResNet-101, ShuffleNet, and MobileNet-v2, are combined using majority voting for better classification performance. The method achieves a maximum classification accuracy of 96.54% for quality images and 95.58% for noisy images, outperforming existing methods, making it a valuable tool for precise diagnosis of kidney diseases. Pirmoradi et al. [28] proposes a new machine learning approach for identifying significant miRNAs and classifying kidney cancer subtypes to develop an automatic diagnostic tool. The method involves two main steps: feature selection using the AMGM measure to choose candidate miRNAs and classification using a self-organizing deep neuro-fuzzy system, which overcomes challenges in high-dimensional data analysis. The results shows that the proposed method achieves high accuracy in classifying kidney cancer subtypes based on the selected miRNAs. Abdeltawab et al. [29] proposes a deep learning pipeline for automated classification of kidney cancer subtypes, specifically clear cell renal cell carcinoma and clear cell papillary renal cell carcinoma. The model uses convolutional neural networks on whole slide images divided into patches, providing patchwise and pixelwise classification. The approach accurately classifies the four classes and outperforms other state-of-the-art methods. This deep learning method has the potential to assist pathologists in diagnosing kidney cancer subtypes from histopathological images. Abdeltawab et al. [30] presents a deep learning framework for classifying kidney tumor subtypes, specifically clear cell renal cell carcinoma and clear cell papillary renal cell carcinoma. The framework utilizes three convolutional neural networks to process kidney image patches of different sizes, providing patchwise and pixelwise classification. The results demonstrate superior performance compared to existing methods, highlighting the potential of deep learning techniques in cancer diagnosis. Khan et al. [31] addressed urgent brain tumor diagnosis using automated methods, specifically convolutional neural networks (CNNs). Deep convolutional features improve classification accuracy significantly, and an ensemble of XGBoost, AdaBoost, and Random Forest achieves a top accuracy of 95.9% for tumors and 94.9% for normal cases, surpassing individual methods. Zhu et al. [32] presents a pipeline employing transfer learning

to address the limitations of small medical image datasets in deep learning applications. The proposed dual-channel fine segmentation network (FS-Net) effectively segmented kidney and tumor regions in 3D CT images, outperforming state-of-the-art methods. The classification model using radiomics features demonstrated accurate classification of benign and malignant tumors in the small dataset. The work emphasizes the significance of architecture design in transfer learning and provides valuable insights for small data analysis in medical imaging. Gulzar et al. [33] conducted using CT scans of 125 subjects, reveals a negative correlation between visceral fat to abdomen size ratio and mean liver intensity values, as well as between mean liver intensity values and total abdomen fat to abdomen size ratio. These correlations indicate a direct link between obesity and diffuse liver fat. This insight contributes to understanding fatty liver disease and its associated health risks. Sarada et al. [34] proposes a hybrid ensemble of visual capsule networks and deep feed-forward extreme learning machines for kidney tumor classification and segmentation from CT images. The model achieves a DICE coefficient of 0.96 and an accuracy of 97.5%, outperforming other hybrid deep learning models. Zhao et al. [35] purpose of the study was to develop an automated method using 3D U-Net and ResNet for accurate segmentation and classification of renal masses in CT images. The algorithm achieved high performance in kidney boundary segmentation (Dice coefficient of 0.99) and renal mass delineation (average Dice coefficients of 0.75 and 0.83). The classification accuracy for masses was 86.05% for masses <5 mm and 91.97% for masses \geq 5 mm. The proposed method demonstrated the capability of accurately localizing and classifying renal masses.

In this study, proposes a computer-assisted diagnosis system using machine learning and fine-tuned transfer learning for efficient kidney tumor detection in CT images, achieving high accuracy using different deep learning models. The contribution of this study is as follows:

- **Enhanced Kidney Tumor Detection:** The study introduces a computer-assisted diagnosis (CAD) system utilizing machine learning and fine-tuned transfer learning, which improves the accuracy and efficiency of kidney tumor detection in CT images compared to manual methods.
- **ML Model Evaluation:** The research comprehensively experiments with various machine learning models, including Random Forest, Support Vector Machine, Gradient Boosting, and Light Gradient Boosting Model, to identify kidney tumors in CT images. The comparison of these models helps identify the most effective approach for tumor detection.
- **Fine-tuned Transfer Learning:** The study proposes and evaluates the performance of fine-tuned ResNet-101 and DenseNet-121 models for computer-assisted diagnosis of kidney tumors. These models demonstrate superior accuracy, precision, recall, and F1-score compared to other models and state-of-the-art methods.

Efficient Data Preprocessing: The research applies different pre-processing techniques and image resizing to reduce the complexity of the training model and speed up the

training process. This optimization ensures faster and more efficient kidney tumor classification.

II. MATERIAL AND METHODS

This section provides an elaborate exposition of the proposed framework, incorporating established machine learning (ML) algorithms and deep learning models using transfer learning (TL). As depicted in Fig. 1, the framework encompasses two core components: data pre-processing and the scaling of input image dimensions. Subsequently, a diverse set of ML and TL models were trained on the dataset, comprising Chest Computed Tomography (CT) images.

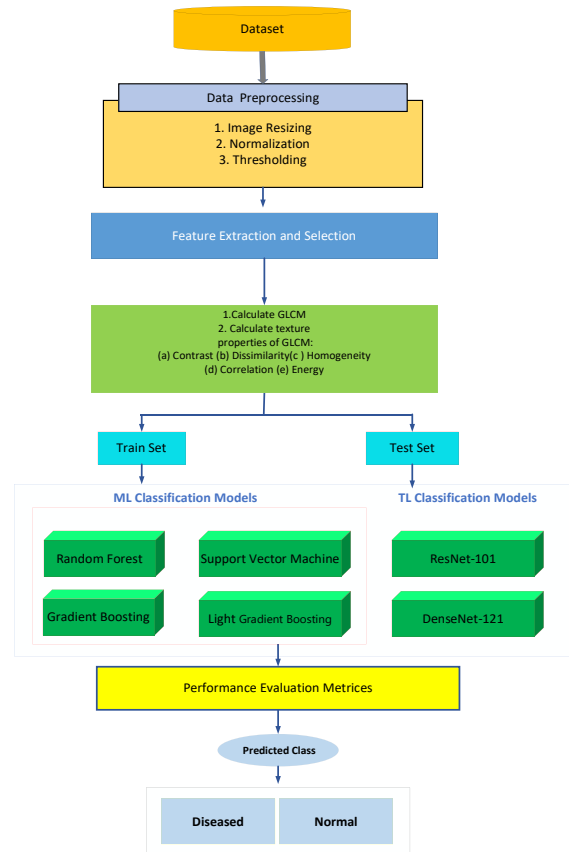


Fig. 1. Detailed framework for proposed methodology.

A. Dataset Description

The research study was conducted using a dataset of CT Kidney images, sourced from an official repository accessible [36]. This dataset was compiled from various hospitals in Bangladesh and Dhaka, obtained through the Picture Archiving and Communication System (PACS). The dataset comprises a total of 12,446 images, distributed across different categories: 3,709 images representing cyst, 5,077 images of normal kidney, 1,377 images of kidney stone, and 2,283 images of kidney tumor patients. Each image in the dataset possesses a resolution of 512 x 512 pixels.

The research investigation primarily focused on two classes within the dataset, specifically the normal and tumor images. Fig. 2 present the CT slide of normal and tumorous kidney samples. To facilitate the experimentation process, the dataset was stratified into three distinct subsets, maintaining a

partitioning ratio of 6:2:2, representing 60% for training, 20% for validation, and 20% for testing purposes, respectively. Such a division ensures a robust evaluation of the proposed methods and allows for the assessment of model performance across distinct data subsets.

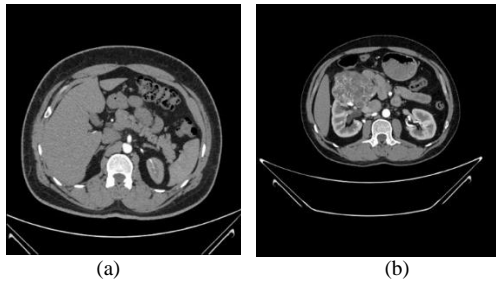


Fig. 2. (a) CT slice of a normal kidney (b) CT slice of kidney with tumor.

1) *Data Pre-processing*: This dataset is taken through three distinct data pre-processing stages before being experimented as discussed below:

a) *Thresholding*: It is the simplest type of image segmentation technique which is applied on the gray scale images to convert them to binary images. The coloured images are converted to binary images using thresholding [37]. Thresholding is experimented to divide a picture into less sections, or trashes, utilizing somewhere around one shading or dim scale an incentive to characterize their boundaries. The complexity of the information is simplified, and the process of detecting and characterising the data can be clarified if a binary image is obtained first is one of the advantages of this approach. Selecting a single Threshold value (T) is the method that is the most well-known for transforming a grayscale image into a binary one. Mathematically it can be interpreted as follows:

$$o_{ij} = \begin{cases} 1 & \text{if } I_{ij} \geq T \\ 0 & \text{if } I_{ij} < T \end{cases} \quad (1)$$

b) *Data normalization*: The image pixels are normalized to a single scale of 0 to 255 gray scale values.

c) *Image resizing*: The given dataset images of dimension 512 x 512 are resized to (32x 32), (64 x 64), and (128x 128) to perform the experiment.

2) *Feature engineering*: It is the amalgamation of the feature extraction, selection and matrix creation. In the proposed research work, features are extracted from the given input images. The Grey Level Co-occurrence matrix (GLCM) [38] has been generated for all the images for the extraction of the relevant features. The GLCM is calculated to analyse the texture of the given image based upon the contrast, homogeneity, Correlation between the matrix, dissimilarity and energy of the pixels. The GLCM uses the distance(d) and angle (θ) between pair of pixels (i, j) to calculate the occurrence of the (i, j)th pair in the image in the direction of the given angle value. The θ value varies from 0 to 360 degree. The size of the matrix depends upon the number of pixel intensities present in the given image for which the matrix has to be calculated.

For example: If there are four intensity values present in the image, then the size of the GLCM will be 4 x 4. The GLCM is calculated for the three different image dimensions i.e. (32x 32), (64 x 64), and (128x 128). But different image sizes do not have much impact on the classification accuracy. So we consider the smallest image size for the feature extraction and model training in order to avoid the unnecessary computation cost. The process of GLCM is repeated for four times. The final extracted features are then combined and given to the Light Gradient Boosting Machine (LGBM) for training.

B. Machine Learning Methods

1) *Random Forest (RF)*: The RF [39] algorithm is a well-known ensemble learning method that is utilized for classification purpose, including kidney disease classification. Random Forest is applied on kidney disease dataset which is prepared with features and corresponding target labels. The features may include various clinical and demographic factors related to kidney disease, while the target labels indicate the presence or absence of kidney disease.

Random Forest Model training follows the below mentioned steps:

- **Tree Construction**: RF consists of multiple decision trees (DT). Where each tree is constructed during training by taking a random subset from training data, known as bootstrap samples. This sampling technique introduces variation in the training process.
- **Feature Sampling**: An unsystematic subset of features is considered at each node of the decision tree, for splitting known as feature sampling or feature bagging. It helps in reducing correlation among the trees and promotes diversity in the ensemble.
- **Splitting Criteria**: The decision tree nodes are split using a splitting (Gini impurity) criterion that helps determining the best feature and threshold to divide the data at each node, aiming to maximize the separation of the target classes.
- **Tree Growth**: The trees are grown until a stopping condition is met. Controlling the tree growth helps prevent overfitting and promotes generalization.

a) *Voting Mechanism*: Once the RF model is trained, predictions are made by aggregating the predictions of all the individual decision trees. The final predictions made by each tree in the forest is determined by majority vote obtained independently, and the final prediction is determined by a majority vote or averaging, depending on the task (binary or multi-class classification).

b) *Class Probability Estimation*: RF can also estimate the probability of belonging to each class. It does this by calculating the proportion of trees that predicted each class. The class probabilities can be useful for assessing the confidence of predictions or for other downstream tasks.

The trained RF is evaluated on the testing dataset to assess its performance using metrics that present insights into the

model's prognostic accuracy and capability to correctly classify instances of kidney disease. The trained Random Forest model can be deployed to make predictions on new, unseen data. It takes the relevant features of a patient as input and provides the predicted class or class probabilities for kidney disease. RF offers several advantages for kidney disease classification, including the ability to handle high-dimensional datasets, handle missing values, and provide feature importance measures. It is a robust and widely used algorithm for classification tasks, including medical diagnosis and risk assessment.

2) *Support Vector Machine (SVM)*: (SVM) [40] is supervised learning algorithm commonly used for classification tasks. The internal architecture of an SVM involves several key components and steps.

- **Relevant features**: Relevant features from the kidney disease dataset are extracted. These features could include measurements such as blood pressure, serum creatinine levels, age, etc.
- **Data Normalization**: The retrieved characteristics are then normalized such that they are all on a scale that is comparable to one another. This step helps in preventing definite features from dominating the others due to their larger magnitudes.
- **Feature Representation**: Each instance or sample of dataset is presented as a feature vector, where each feature corresponds to a specific attribute or measurement related to kidney disease.
- **Hyperplane Initialization**: In a high-dimensional feature space, the SVM finds the best hyper plane to split data points by class. This hyperplane gets configured at the outset by the SVM algorithm.
- **Training**: The SVM algorithm trains the model by iteratively optimizing the position and orientation of the hyperplane to get the most out of the margin between the classes. The goal is to find the hyperplane that separates the classes by minimizing the misclassification. The optimization problem is typically solved using techniques such as the Sequential Minimal Optimization (SMO) algorithm or other quadratic programming methods. During the training process, subset of data points called support vectors are identified that are closest to the hyperplane or located inside the boundary.
- **Classification**: After training is complete, SVM is used for classification of new, unseen instances. The decision boundary of the SVM is defined by the hyperplane, and the side of the hyperplane on which a data point lies determines its predicted class label. The distance of a data point from the hyperplane can also provide additional information about the model's confidence in its prediction.

3) *Gradient Boosting (GB)*: GB [41] is an ensemble learning method that incorporated multiple weak learners, to

create a strong prognostic model. GB is used for kidney disease classification is discussed stepwise as below:

- **Data Preparation**: Similar to other machine learning algorithms, the kidney disease dataset needs to be prepared by extracting relevant features and normalizing the data if necessary.
- **Initialization**: Gradient Boosting starts with an initial model, often a simple one like a decision tree with limited depth (weak learner). The initial model is typically assigned equal weights for all samples in the training set.
- **Iterative Training**: In each iteration, a new weak learner, referred to as a "base learner," is trained to approve the miss-classifications made by the ensemble of models trained so far. The base learner is fitted to the training set, with a focus on the samples that were misclassified or had high residuals from the previous iteration. The learning process involves minimizing a loss function. The base learner is typically a decision tree that is grown using a greedy algorithm, selecting the best split points based on information gain or other criteria. The depth and complexity of the decision tree can be adjusted to balance model performance and computational efficiency.
- **Boosting and Weight Updates**: After training the base learner, its predictions are combined with the predictions of the previous models in the ensemble. Initially, all models are given equal weights. However, the subsequent models focus on the samples that were misclassified or had high residuals from the previous models. The weights of the samples are adjusted to prioritize the challenging instances. The misclassified samples are assigned higher weights, while correctly classified samples have lower weights. This process emphasizes the samples that are difficult to classify, allowing subsequent models to concentrate on improving their predictions.
- **Iteration and Ensemble Building**: Steps 3 and 4 are repetitive for a predetermined number of iterations or until a certain performance threshold is reached. In each iteration, a new base learner is trained to minimize the weighted loss function. The predictions of all models in the ensemble are combined using a weighted sum or averaging scheme, where the weights are determined by the performance of each model.
- **Final Prediction**: The final prediction for a new, unseen instance is obtained by aggregating the predictions of all models in the ensemble, typically using a majority vote or weighted average. For classification tasks, the predicted class label is determined based on the aggregated predictions.

4) *Light Gradient Boosting Model (LGBM)*: LGBM [42] is a powerful gradient boosting framework that combines speed and efficiency with high predictive accuracy. Its ability to handle large datasets efficiently and its regularization techniques make it a popular choice among data scientists and

machine learning practitioners. Light GBM is based on the gradient boosting framework, which is an ensemble method. It sequentially trains the models to overcome the drawbacks of the previous models, thus getting better the overall prognostic accuracy.

The kidney disease dataset is prepared with features and corresponding target labels. The features may include various clinical and demographic factors related to kidney disease, such as age, blood pressure, creatinine levels, etc. The target labels indicate the presence or absence of kidney disease. Model Configuration: The LGBM model is configured with the following parameters:

- Learning Rate: The learning rate determines the step size at each boosting iteration. In this case, the learning rate is set to 0.05, indicating a relatively small step size.
- Boost Type: The boost type is set to "Dart," which refers to the Dart boosting algorithm. Dart is a variation of gradient boosting that introduces dropout regularization to prevent overfitting.
- Metric: The chosen evaluation metric is "multi log loss," which is suitable for multi-class classification problems. It calculates the logarithmic loss between the predicted labels and true class labels.
- Number of Leaves: The LGBM model is configured with 100 leaves. The leaves represent the final decision regions of the ensemble of decision trees.
- Max Depth: of each individual decision tree in the ensemble is set to ten. This parameter limits the complexity of the trees and helps prevent overfitting.
- Class: The classification task involves predicting between two classes, likely representing the presence or absence of kidney disease.

The LGBM model is trained on the training dataset using the configured parameters. The model uses gradient boosting to iteratively fit decision trees to the training data, improving its predictive performance at each iteration. Once the training is complete, the trained LGBM model is evaluated on the testing dataset using the multi log loss metric. It can take the relevant features of a patient as input and provide the probability or predicted class of kidney disease. By following this architectural workflow, the LGBM model with the specified parameters can be effectively trained and used for kidney disease classification. It is important to note that further customization and tuning of the parameters might be necessary depending on the specific description of the dataset and the desired performance. The pseudo code for the proposed research methodology is given in Pseudocode 1.

1. Pseudocode for ML models for kidney disease classification

Step 1: Load and preprocess the dataset // Load dataset
Preprocess dataset (handle missing values, encode categorical variables, etc.)
Split dataset into: features (X) and target variable (y)

```
Step 2: train_test_split is done
Step 3: Model1 = RFClassifier(),
      Model2 = GBClassifier(),
      Model3 = SVMClassifier(),
      Define the model "Light GBM" and set hyperparameters
      Model4 = LGBMClassifier( boosting_type='dart',
      objective='binary',
      metric='multi_logloss', num_leaves = <100>,
      learning_rate = <0.05>,
      feature_fraction = <feature_fraction>,
      bagging_fraction = <bagging_fraction>,
      verbose = -1)
Step 4: model.fit(X_train, y_train) //Train the Light GBM model
Step 5: y_pred = model.predict(X_test) //Make predictions and
evaluate the model
Step 6: Accuracy = accuracy_score(y_test, y_pred) // Calculate
evaluation metrics
      CM = confusion_matrix(y_test, y_pred)
Step 7: Tune hyperparameters to optimize model performance
Step 8: Perform further analysis (e.g., feature importance)
```

LGBM has shown excellent prediction for kidney disease classification. Its ability to iteratively refine the model by focusing on the challenging instances makes it effective in capturing complex relationships in the data and improving predictive accuracy.

C. Transfer Learning Architectures

1) *ResNet-101*: ResNet-101 [43] features a 101-layer deep CNN. The weights of the network's pre-trained version on over a million photos from the ImageNet database may be imported. The trained network can classify photos into 1000 different things, such as keyboards, mice, books, and other stuff. To train the model, an Adam optimizer with a learning rate of 4e-5 and binary cross entropy as the loss function was used. During model training, the total number of trainable parameters was 525,313.

2) *DenseNet-121*: It is a dense network [44] of 121 layers, 120 of which are dense layers and four of which are average pool layers. The weights of all layers in the same deep dense block are circulated across the inputs, allowing the deep layers to utilise the early extracted features. DenseNet uses features more efficiently and outperforms with less parameters. The network is trained over ten epochs using an Adam optimizer and a learning rate of 4e-5 hyperparameters. The total number of trainable parameters in the model was 263,169.

The internal architecture for fine-tuned ResNet-101 and DenseNet-121 are given in Fig. 3 and 4, respectively. The accuracy of the model was found to be enhanced when (32, 32) pixel images were used as input. The smaller image size results in a lower computation and time complexity, which may cause the model to learn features or patterns in the images faster. The Adam optimizer was used to train the model since it is the best optimizer for early convergence. When employing Adam as an optimizer, the model learns up new information more quickly. In order to train our suggested model, we implement a learning rate schedule in the Adam. The training of the model lasted 50 epochs, with a batch size of 32, and the initial learning rate was set at 0.001.

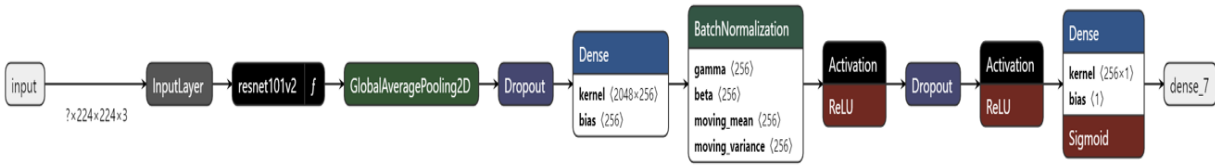


Fig. 3. Architecture of ResNet-101 model for diagnosing KT using CT images.

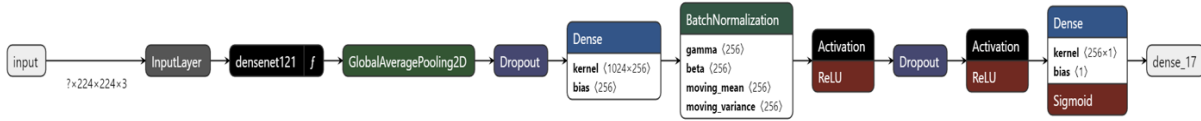


Fig. 4. Architecture of DenseNet-121 model for diagnosing KT using CT images.

D. Experimental Environment Settings and Performance Evaluation Metrics

This research aims to propose an optimal model which identifies and classifies different types of images. The proposed model was implemented using Python (v. 3.8), OpenCV (v. 4.7), Keras Library (v. 2.8) were used on Windows 10 Pro OS, with system configuration using an Intel i5 processor running at 2.9 GHz, an Nvidia RTX 2060 Graphical Processing Unit and 16 GB RAM.

Several metrics were employed to evaluate the performance of classifying sunflower blooms and leaves, including accuracy, precision, recall, and F1-score, which are frequently used indicators [45]. Accuracy is the ratio of samples from all classes that can be correctly identified, Recall is the ratio of correctly classified positives among all actual positives, and Precision is the ratio of correctly identified positives versus all expected positives [46]. The metrics were calculated using Eq. (1) - (4).

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

$$Precision = \frac{TP}{True\ Positive + False\ Positive} \quad (3)$$

$$F1 - Score = 2x \frac{Recall \times Precision}{Recall + Precision} \quad (4)$$

$$AUC = \sum_{i=0}^N \left(\frac{TPI(i) + TPI(i-1)}{2} \right) (FPI(i) - (FPI(i-1))) \quad (5)$$

III. RESULTS AND DISCUSSION

A. Performance Comparison

The ML models were meticulously trained on a dataset, and to bolster their performance, advanced feature extraction using the GLCM technique was employed. This technique allowed the models to capture intricate patterns and textures from the kidney images, enhancing their ability to predict Kidney Tumors (KT) more accurately. To push the boundaries of prediction accuracy even further, two state-of-the-art architectures, ResNet-101 and DenseNet-121, were extensively experimented with. These cutting-edge models were chosen for their exceptional deep learning capabilities,

enabling them to unravel complex relationships within the data and make precise predictions regarding the presence of kidney tumors.

Following the intensive training process, the models' prowess was thoroughly evaluated using a diverse range of performance metrics, such as accuracy, recall, precision, and F1_score. These metrics provided a comprehensive view of the models' overall classification performance and offered valuable insights into their strengths and potential areas of improvement. In addition to the conventional performance metrics, the proposed model's efficacy was also measured using the Area Under the Curve (AUC) on the test set. AUC is a significant indicator of the model's ability to distinguish between positive and negative cases, providing a more comprehensive understanding of its discriminative power.

The outcomes of the classification process were further analyzed to gain a deeper understanding of the models' predictions. Specifically, the binary class results (class 0 and class 1) were meticulously generated and assessed. Label 0 corresponded to the classification results for normal kidneys, indicating instances where the model correctly identified normal kidneys (True negative). On the other hand, Label 1 represented the classification results for tumorous kidneys, showcasing the model's capacity to accurately detect kidney tumors (True positive).

Moreover, the results also provided insights into false positives and false negatives, highlighting instances where the model might have made errors in its predictions. This comprehensive analysis aimed to identify potential areas of improvement and guide future iterations of the models. By combining advanced feature extraction techniques, cutting-edge deep learning architectures, and a thorough evaluation using diverse performance metrics, this study aimed to achieve a robust and reliable prediction model for Kidney Tumor detection, which could have a significant impact on the early diagnosis and treatment of kidney diseases.

Table I presents the results of various performance metrics calculated on the Test set for different machine learning algorithms, as well as for the proposed fine-tuned Transfer Learning architectures, ResNet-101, and DenseNet-121. The metrics evaluated are Accuracy, Precision, Recall, F1_score, and AUC.

TABLE I. RESULTS OF THE DIFFERENT PERFORMANCE METRICS CALCULATED ON THE TEST SET

Model	Accuracy	Precision	Recall	F1_score	AUC
LGBM	94.09	0.9510	0.9352	0.9595	0.9552
GB	92.44	0.9439	0.8991	0.9843	0.9773
SVM	91.27	0.9300	0.8890	0.8906	0.8834
RF	91.02	0.9201	0.9212	0.9656	0.9588
ResNet-101	96.67	0.9532	0.9111	0.9843	0.9773
DenseNet-121	98.22	0.9577	0.9323	0.9843	0.9773

From the table it can be noticed that both ResNet-101 and DenseNet-121 achieved the highest accuracy among all the models. DenseNet-121 outperformed all other models with an impressive accuracy of 98.22%, while ResNet-101 achieved an accuracy of 96.67%. The success of these Transfer Learning architectures can be attributed to their pre-trained weights and knowledge gained from large datasets, allowing them to recognize and learn intricate patterns and features from the given kidney tumor dataset effectively. LGBM and GB achieved reasonably high accuracies of 94.09% and 92.44%, respectively. LGBM performed slightly better than GB, which indicates the effectiveness of gradient boosting in ensemble learning. However, the accuracies of both LGBM and GB were lower compared to the Transfer Learning models. SVM and RF achieved accuracies of 91.27% and 91.02%, respectively. While SVM relies on finding optimal hyperplanes for classification, RF uses an ensemble of decision trees. Although these algorithms achieved respectable accuracies, they were outperformed by the Transfer Learning models. The precision, recall, and F1_score metrics provide insights into the models' ability to correctly classify positive and negative instances, as well as their overall predictive performance. DenseNet-121 consistently achieved the highest F1_score of 0.9843, indicating its superior balance between precision and recall in classifying both tumor and normal kidney instances. AUC is a measure of the models' ability to distinguish between positive and negative cases. Remarkably, both ResNet-101 and DenseNet-121 attained an AUC of 0.9773, matching the performance of GB. This demonstrates the Transfer Learning models' robustness in making accurate predictions and differentiating between tumor and normal kidney instances.

Accuracy is a crucial metric in evaluating the performance of machine learning models, as it represents the percentage of correct predictions made by the model where the predicted value aligns with the real value. Throughout the training phase, accuracy is continuously monitored and plotted, providing valuable insights into the model's learning progress and its ability to make accurate predictions as it iteratively updates its weights and biases.

While accuracy provides an overall view of the model's correctness, it is essential to delve deeper into the model's learning dynamics. For this purpose, loss functions play a pivotal role in assessing the model's performance. Loss functions measure the disparities between predicted values and actual ground-truth values, quantifying the uncertainty or error in the model's estimates. By optimizing the loss during training, the model learns to minimize the discrepancies and improve its predictive capability.

In Fig. 5 and Fig. 6, we present the accuracy and loss plots for two cutting-edge deep learning architectures, ResNet-101 and DenseNet-121, respectively. These plots showcase how accuracy improves and loss decreases over the training iterations, providing a comprehensive view of the models' learning behaviors. The ascending accuracy curve demonstrates how the models become more adept at correctly classifying kidney tumor images as training progresses. Simultaneously, the descending loss curve indicates that the models effectively minimize prediction errors, leading to more precise and confident predictions. Analyzing the accuracy and loss plots for ResNet-101 and DenseNet-121 offers valuable insights into their learning dynamics and convergence patterns. These visualizations not only validate the models' effectiveness in the classification task but also aid in fine-tuning the hyperparameters or adjusting the training strategy to achieve optimal performance. By carefully monitoring the accuracy and loss during training, we gain a deeper understanding of the models' efficacy and can confidently assert that both ResNet-101 and DenseNet-121 exhibit exceptional learning capabilities, making them powerful tools for kidney tumor classification.

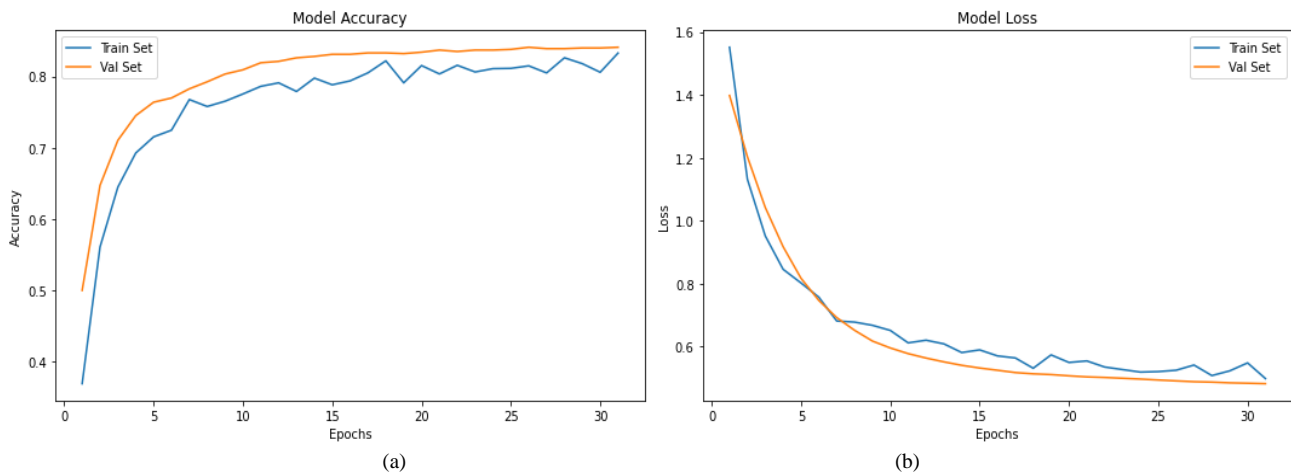


Fig. 5. Graphs plotted between model (a) accuracy and (b) loss over no of epochs for ResNet-101.

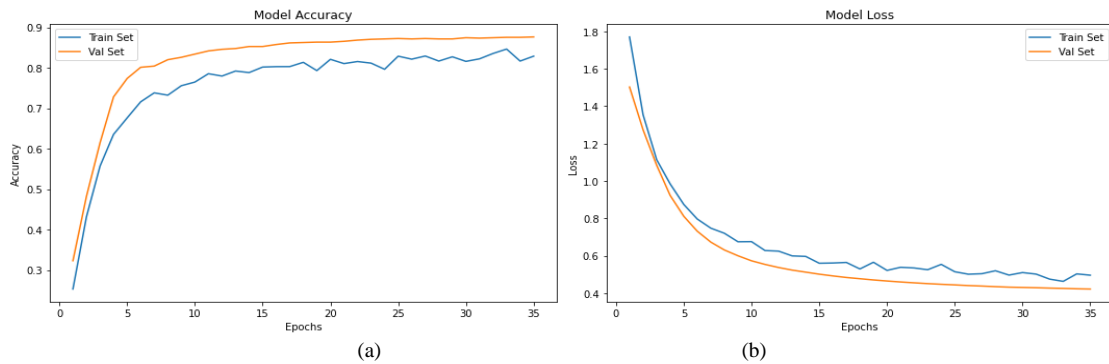


Fig. 6. Graphs plotted between model (a) accuracy and (b) loss over no of epochs DenseNet-121.

In addition to accuracy and loss analysis, the confusion matrix (CM) stands as a pivotal tool for the comprehensive evaluation of classification models, such as ResNet-101 and DenseNet-121. The CM provides valuable insights into the model's performance by breaking down the predictions into four fundamental categories: true positives, true negatives, false positives, and false negatives. It highlights the model's ability to correctly classify positive and negative instances, as well as its potential for making erroneous predictions.

Fig. 7(a) and 7(b) depict the confusion matrices specifically for ResNet-101 and DenseNet-121, respectively. These visual representations offer a detailed view of the models' classification outcomes, enabling us to gauge their effectiveness in distinguishing between normal and tumorous kidney images.

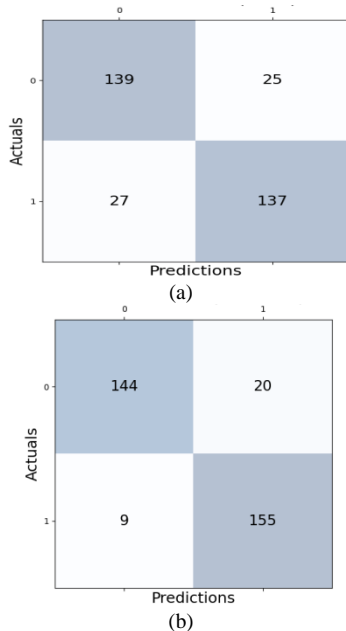


Fig. 7. Confusion Matrix of (a) ResNet-101 and (b) DenseNet-121.

B. K-Fold Cross Validation

K-fold cross-validation is a widely used technique in machine learning to assess the performance and generalization capability of a model while mitigating the risks of overfitting. Overfitting occurs when a model performs well on the training data but fails to generalize to unseen data, which can lead to

inflated evaluation metrics on the test set. To address this concern, the model's performance is tested using a separate validation set, and 10-fold cross-validation is a popular approach to achieve this.

In 10-fold cross-validation, the dataset is divided into 10 subsets of approximately equal size. The model is then trained and evaluated ten times, each time using a different subset as the validation set and the remaining nine subsets for training. This ensures that the model is evaluated on different partitions of the data, providing a more robust estimate of its performance and reducing the influence of any particular data split.

Table II presents the results of 10-fold cross-validation on the CT Kidney image dataset for different machine learning (ML) models and fine-tuned Transfer Learning (TL) models, namely Random Forest, Support Vector Machine (SVM), Gradient Boost, LGBM (LightGBM), ResNet-101, and DenseNet-121. The accuracies achieved by each model on each fold are shown.

Observations from the 10-fold Cross Validation Results:

1) Random Forest: Random Forest demonstrates relatively stable performance across the folds, with accuracy ranging from 70.12% to 84.21%. Its mean accuracy is 77.09%.

2) Support Vector Machine (SVM): SVM shows higher accuracy values, with a range of 66.23% to 81.81% across the folds. The mean accuracy is 88.82%, making it one of the best-performing ML models in this study.

3) Gradient Boost: Gradient Boosting performs consistently well, with accuracy varying between 62.33% to 79.12%. The mean accuracy achieved by Gradient Boost is 89.11%.

4) LGBM (LightGBM): LGBM outperforms other ML models, exhibiting accuracy in the range of 89.12% to 93.32%. Its mean accuracy is 90.71%.

5) ResNet-101 and DenseNet-121 (Transfer Learning): Both fine-tuned Transfer Learning models, ResNet-101, and DenseNet-121 consistently achieve higher accuracy values compared to ML models. ResNet-101 achieves accuracy between 86.72% to 96.81%, with a mean accuracy of 91.61%. DenseNet-121 exhibits even better performance, with accuracy ranging from 90.0% to 97.61% and a mean accuracy of 92.44%.

TABLE II. 10 FOLD CROSS VALIDATION PERFORMANCE OF THE ML AND FINE-TUNED TL MODELS ON THE CT KIDNEY IMAGE DATASET

10-folds	Random Forest	Support Vector Machine	Gradient Boost	LGBM	ResNet-101	DensNet-121
Fold-1	79.22	81.81	74.02	91.62	88.92	96.81
Fold-2	70.12	66.23	70.12	89.12	86.72	92.02
Fold-3	80.51	79.22	72.83	88.92	96.62	94.62
Fold-4	79.22	75.32	79.12	89.23	92.62	94.82
Fold-5	72.72	71.42	77.42	91.92	91.02	91.72
Fold-6	81.81	77.92	79.12	92.32	90.02	92.82
Fold-7	71.42	68.83	62.33	93.23	92.33	87.32
Fold-8	75.32	76.62	78.42	91.83	91.62	90.32
Fold-9	76.31	76.31	71.05	93.32	90.0	91.71
Fold-10	84.21	81.57	77.63	89.57	90.26	97.61
Mean	77.09	88.82	89.11	90.71	91.61	92.44

Overall, the results demonstrate the effectiveness of the fine-tuned Transfer Learning models, ResNet-101 and DenseNet-121, in accurately classifying kidney images. These models outperform the traditional ML algorithms, such as Random Forest, SVM, and Gradient Boost, in terms of accuracy. The use of 10-fold cross-validation provides a more reliable estimate of the models' performance and their generalization capability, ensuring that the evaluations are robust and less affected by variations in the data splits.

In conclusion, the 10-fold cross-validation results reveal the superior performance of the fine-tuned Transfer Learning models, ResNet-101, and DenseNet-121, in accurately classifying CT kidney images. These models are better equipped to handle the complexities of the dataset, offering promising implications for kidney tumor detection and diagnosis in a real-world clinical setting.

Furthermore, Table III presents a comparison of the results obtained using LightGBM (LGBM) with state-of-the-art methods in terms of accuracy for a specific task or dataset. The table showcases the performance of various models, including two fine-tuned Transfer Learning architectures, ResNet-101 and DenseNet-121, along with several other approaches reported in the literature.

Table III demonstrate the effectiveness of the fine-tuned Transfer Learning models, ResNet-101 and DenseNet-121, in achieving high accuracies and outperforming other state-of-the-art methods. DenseNet-121, in particular, exhibits the highest accuracy among all the models, indicating its superior performance in the specific classification task. The results also highlight the importance of exploring and comparing different models and methodologies to advance the field and achieve better results in image classification and other related tasks.

TABLE III. COMPARISON OF PROPOSED WORK WITH STATE-OF-THE-ART METHODS

Reference	Accuracy (%)
ResNet-101	96.67
DenseNet-121	98.22
LGBM	94.09
Zhou et al.[47]	93.00
Zabihollahy et al. [48]	83.75
Schieda et al. [49]	78.00
Finally, Yap et al. [50]	75.00

IV. CONCLUSION

The early detection and classification of kidney tumors play a vital role in saving human lives. Manual detection methods rely heavily on the expertise of medical professionals and can be time-consuming. Therefore, the development of automatic classification systems holds significant promise, as they offer robust and rapid results. In this study, we have presented a hybrid approach combining Light Gradient Boosting Method with Grey Level Co-occurrence matrix (GLCM) computation for the automatic classification of kidney tumors from CT Kidney image datasets. To optimize the training process, we applied various pre-processing techniques and image resizing, reducing the model's complexity and speeding up the training. Light GBM, known for its speed, efficiency, and high predictive accuracy among ML models, served as a powerful gradient boosting framework in this study. Additionally, we introduced two fine-tuned Transfer Learning (TL) models within this framework, ResNet-101 and DenseNet-121, to predict kidney tumors. The performance of these models was thoroughly evaluated using diverse performance metrics and compared with state-of-the-art methods. Our results demonstrated the superiority of the fine-tuned TL models, with DenseNet-121 achieving an impressive accuracy of 98.22%.

Several limitations are evident in the study's comparative analysis, model testing, and generalizability. Firstly, the comparative study of detection approaches, including Random Forest, Support Vector Machine, Gradient Boosting, Light Gradient Boosting Model, and deep learning models ResNet-101 and DenseNet-121, might lack a comprehensive exploration of other relevant models, potentially missing out on valuable insights and alternative solutions. Secondly, while the fine-tuned deep learning models exhibit impressive accuracy, their testing and evaluation solely on the provided dataset might not guarantee similar performance on diverse and real-world datasets. There's a need to assess the models across various datasets to ascertain their consistency and robustness. Finally, the proposed approach's applicability to different datasets with varying characteristics, like imaging quality and patient demographics, remains unexplored. Testing the models on multiple datasets could reveal potential challenges in generalizing the approach to broader clinical settings.

In light of these limitations, future work should aim to address these areas to enhance the study's comprehensiveness and practicality. To conduct a more thorough comparative study, incorporating a wider range of detection models, including emerging techniques and architectures, would provide a more comprehensive understanding of the strengths and weaknesses of various approaches. Moreover, the robustness and generalizability of the fine-tuned deep learning models, ResNet-101 and DenseNet-121, should be evaluated across multiple datasets to ensure consistent performance across diverse clinical scenarios. Additionally, to enhance the proposed approach's real-world applicability, further investigation on different datasets, encompassing variations in imaging quality, patient populations, and demographics, is crucial. This analysis will shed light on potential challenges and adaptations needed to deploy the model effectively in clinical practice. Ultimately, addressing these avenues for future research will contribute to a more holistic and adaptable approach for kidney tumor detection, ensuring its utility and effectiveness across a broader spectrum of clinical settings.

ACKNOWLEDGMENT

King Faisal University: Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under the Project GRANT3, 926.

REFERENCES

- [1] Moch, H.; Cubilla, A.L.; Humphrey, P.A.; Reuter, V.E.; Ulbright, T.M. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *Eur Urol* 2016, 70, doi:10.1016/j.eururo.2016.02.029.
- [2] Rowe, S.P.; Pomper, M.G. Molecular Imaging in Oncology: Current Impact and Future Directions. *CA Cancer J Clin* 2022, 72, 333–352, doi:10.3322/caac.21713.
- [3] Ansari, K.K.; Jha, A. Causes of Cancer in the World: Comparative Risk Assessment of Nine Behavioral and Environmental Risk Factors. *Cureus* 2022, 14, e28875–e28875, doi:10.7759/cureus.28875.
- [4] Kidney Cancer: Statistics Available online: Kidney Cancer: Statistics (accessed on 23 July 2023).
- [5] Khan, Y.F.; Kaushik, B. Neuroimaging (Anatomical MRI)-Based Classification of Alzheimer's Diseases and Mild Cognitive Impairment Using Convolution Neural Network. In *Lecture Notes on Data Engineering and Communications Technologies*; 2022; Vol. 106.
- [6] Irigaray, P.; Newby, J.A.; Clapp, R.; Hardell, L.; Howard, V.; Montagnier, L.; Epstein, S.; Belpomme, D. Lifestyle-Related Factors and Environmental Agents Causing Cancer: An Overview. *Biomedicine & Pharmacotherapy* 2007, 61, 640–658, doi:10.1016/j.biopha.2007.10.006.
- [7] Kocak, B.; Durmaz, E.S.; Kaya, O.K.; Ates, E.; Kilickesmez, O. Reliability of Single-Slice-Based 2D CT Texture Analysis of Renal Masses: Influence of Intra- and Interobserver Manual Segmentation Variability on Radiomic Feature Reproducibility. *American Journal of Roentgenology* 2019, 213, 377–383, doi:10.2214/ajr.19.21212.
- [8] Gulzar, Y. Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique. *Sustainability* 2023, 15, 1906.
- [9] Dhiman, P.; Kaur, A.; Balasaraswathi, V.R.; Gulzar, Y.; Alwan, A.A.; Hamid, Y. Image Acquisition, Preprocessing and Classification of Citrus Fruit Diseases: A Systematic Literature Review. *Sustainability* 2023, Vol. 15, Page 9643 2023, 15, 9643, doi:10.3390/SU15129643.
- [10] Mamat, N.; Othman, M.F.; Abdulghafor, R.; Alwan, A.A.; Gulzar, Y. Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability* 2023, 15, 901.
- [11] Gulzar, Y.; Ünal, Z.; Akta, s, H.A.; Mir, M.S. Harnessing the Power of Transfer Learning in Sunflower Disease Detection: A Comparative Study. *Agriculture* 2023, Vol. 13, Page 1479 2023, 13, 1479, doi:10.3390/AGRICULTURE13081479.
- [12] Aggarwal, S.; Gupta, S.; Gupta, D.; Gulzar, Y.; Juneja, S.; Alwan, A.A.; Nauman, A. An Artificial Intelligence-Based Stacked Ensemble Approach for Prediction of Protein Subcellular Localization in Confocal Microscopy Images. *Sustainability* 2023, Vol. 15, Page 1695 2023, 15, 1695, doi:10.3390/SU15021695.
- [13] Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. *Symmetry (Basel)* 2020, 12, 2018.
- [14] Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. *International Journal on Perceptive and Cognitive Computing* 2021, 7, 85–91.
- [15] Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. *International Journal of Information Technology* 2022 2022, 1–11, doi:10.1007/S41870-022-01130-5.
- [16] Gulzar, Y.; Alwan, A.A.; Abdullah, R.M.; Abualkishik, A.Z.; Oumrani, M. OCA: Ordered Clustering-Based Algorithm for E-Commerce Recommendation System. *Sustainability* 2023, Vol. 15, Page 2947 2023, 15, 2947, doi:10.3390/SU15042947.
- [17] Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. *Applied Sciences* 2022, Vol. 12, Page 5990 2022, 12, 5990, doi:10.3390/APP12125990.
- [18] Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. *Symmetry (Basel)* 2021, 13, 1987.
- [19] Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. *Wirel Commun Mob Comput* 2022, 2022.
- [20] Mehmood, A.; Gulzar, Y.; Ilyas, Q.M.; Jabbari, A.; Ahmad, M.; Iqbal, S. SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions. *Cancers* 2023, Vol. 15, Page 3604 2023, 15, 3604, doi:10.3390/CANCERS15143604.
- [21] Anand, V.; Gupta, S.; Gupta, D.; Gulzar, Y.; Xin, Q.; Juneja, S.; Shah, A.; Shaikh, A. Weighted Average Ensemble Deep Learning Model for Stratification of Brain Tumor in MRI Images. *Diagnostics* 2023, Vol. 13, Page 1320 2023, 13, 1320, doi:10.3390/DIAGNOSTICS13071320.
- [22] Lee, H.; Hong, H.; Kim, J.; Jung, D.C. Deep Feature Classification of Angiomyolipoma without Visible Fat and Renal Cell Carcinoma in Abdominal Contrast-Enhanced CT Images with Texture Image Patches and Hand-Crafted Feature Concatenation. *Med Phys* 2018, 45, 1550–1561, doi:10.1002/mp.12828.
- [23] Han, S.; Hwang, S.I.; Lee, H.J. The Classification of Renal Cancer in 3-Phase CT Images Using a Deep Learning Method. *J Digit Imaging* 2019, 32, 638–643, doi:10.1007/s10278-019-00230-2.
- [24] Tabibu, S.; Vinod, P.K.; Jawahar, C.V. Pan-Renal Cell Carcinoma Classification and Survival Prediction from Histopathology Images Using Deep Learning. *Sci Rep* 2019, 9, doi:10.1038/s41598-019-46718-3.
- [25] Oberai, A.; Varghese, B.; Cen, S.; Angelini, T.; Hwang, D.; Gill, I.; Aron, M.; Lau, C.; Duddalwar, V. Deep Learning Based Classification of Solid Lipid-Poor Contrast Enhancing Renal Masses Using Contrast Enhanced CT. *British Journal of Radiology* 2020, 93, doi:10.1259/bjr.20200002.
- [26] Pedersen, M.; Andersen, M.B.; Christiansen, H.; Azawi, N.H. Classification of Renal Tumour Using Convolutional Neural Networks to Detect Oncocytoma. *Eur J Radiol* 2020, 133, doi:10.1016/j.ejrad.2020.109343.
- [27] Sudharson, S.; Kokil, P. An Ensemble of Deep Neural Networks for Kidney Ultrasound Image Classification. *Comput Methods Programs Biomed* 2020, 197, doi:10.1016/j.cmpb.2020.105709.
- [28] Pirmoradi, S.; Teshnehlab, M.; Zarghami, N.; Sharifi, A. A Self-Organizing Deep Neuro-Fuzzy System Approach for Classification of Kidney Cancer Subtypes Using MiRNA Genomics Data. *Comput*

- Methods Programs Biomed 2021, 206, doi:10.1016/j.cmpb.2021.106132.
- [29] Abdeltawab, H.; Khalifa, F.; Mohammed, M.; Cheng, L.; Gondim, D.; El-Baz, A. A Pyramidal Deep Learning Pipeline for Kidney Whole-Slide Histology Images Classification. *Sci Rep* 2021, 11, doi:10.1038/s41598-021-99735-6.
- [30] Abdeltawab, H.A.; Khalifa, F.A.; Ghazal, M.A.; Cheng, L.; El-Baz, A.S.; Gondim, D.D. A Deep Learning Framework for Automated Classification of Histopathological Kidney Whole-Slide Images. *J Pathol Inform* 2022, 13, doi:10.1016/j.jpi.2022.100093.
- [31] Khan, F.; Ayoub, S.; Gulzar, Y.; Majid, M.; Reegu, F.A.; Mir, M.S.; Soomro, A.B.; Elwasila, O. MRI-Based Effective Ensemble Frameworks for Predicting Human Brain Tumor. *Journal of Imaging* 2023, Vol. 9, Page 163 2023, 9, 163, doi:10.3390/JIMAGING9080163.
- [32] Zhu, X.-L.; Shen, H.-B.; Sun, H.; Duan, L.-X.; Xu, Y.-Y. Improving Segmentation and Classification of Renal Tumors in Small Sample 3D CT Images Using Transfer Learning with Convolutional Neural Networks. *Int J Comput Assist Radiol Surg* 2022, 17, 1303–1311, doi:10.1007/s11548-022-02587-2.
- [33] Gulzar, Y.; Alkinani, A.; Alwan, A.A.; Mehmood, A. Abdomen Fat and Liver Segmentation of CT Scan Images for Determining Obesity and Fatty Liver Correlation. *Applied Sciences* 2022, Vol. 12, Page 10334 2022, 12, 10334, doi:10.3390/AP122010334.
- [34] Sarada, J.; Lakshmi, N.V.M.; Praveena, T.L. U-CapKidnets++: A Novel Hybrid Capsule Networks with Optimized Deep Feed Forward Networks for an Effective Classification of Kidney Tumours Using CT Kidney Images. *International Journal on Recent and Innovation Trends in Computing and Communication* 2022, 10, 274–283, doi:10.17762/ijritcc.v10i1s.5849.
- [35] Zhao, T.; Sun, Z.; Guo, Y.; Sun, Y.; Zhang, Y.; Wang, X. Automatic Renal Mass Segmentation and Classification on CT Images Based on 3D U-Net and ResNet Algorithms. *Front Oncol* 2023, 13, doi:10.3389/fonc.2023.1169922.
- [36] T KIDNEY DATASET: Normal-Cyst-Tumor and Stone Available online: <https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone> (accessed on 24 June 2023).
- [37] Sahoo, P.K.; Soltani, S.; Wong, A.K.C. A Survey of Thresholding Techniques. *Comput Vis Graph Image Process* 1988, 41, 233–260, doi:10.1016/0734-189x(88)90022-9.
- [38] V, B.S. Grey Level Co-Occurrence Matrices: Generalisation and Some New Features. *International Journal of Computer Science, Engineering and Information Technology* 2012, 2, 151–157, doi:10.5121/ijcseit.2012.2213.
- [39] Breiman, L. Random Forests. *Mach Learn* 2001, 45, 5–32, doi:10.1023/A:1010933404324/METRICS.
- [40] Vapnik, V.; Golowich, S.E.; Smola, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In *Proceedings of the Advances in Neural Information Processing Systems*; 1997.
- [41] Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat* 2001, 29, doi:10.1214/aos/1013203451.
- [42] Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv Neural Inf Process Syst* 2017, 30.
- [43] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*; 2016; pp. 770–778.
- [44] Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In *Proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition*; 2017; pp. 4700–4708.
- [45] Ayoub, S.; Gulzar, Y.; Reegu, F.A.; Turaev, S. Generating Image Captions Using Bahdanau Attention Mechanism and Transfer Learning. *Symmetry (Basel)* 2022, 14, 2681.
- [46] Ayoub, S.; Gulzar, Y.; Rustamov, J.; Jabbari, A.; Reegu, F.A.; Turaev, S. Adversarial Approaches to Tackle Imbalanced Data in Machine Learning. *Sustainability* 2023, Vol. 15, Page 7097 2023, 15, 7097, doi:10.3390/SU15097097.
- [47] Zhou, L.; Zhang, Z.; Chen, Y.-C.; Zhao, Z.-Y.; Yin, X.-D.; Jiang, H.-B. A Deep Learning-Based Radiomics Model for Differentiating Benign and Malignant Renal Tumors. *Transl Oncol* 2019, 12, 292–300, doi:10.1016/j.tranon.2018.10.012.
- [48] Zabihollahy, F.; Schieda, N.; Krishna, S.; Ukwatta, E. Automated Classification of Solid Renal Masses on Contrast-Enhanced Computed Tomography Images Using Convolutional Neural Network with Decision Fusion. *Eur Radiol* 2020, 30, 5183–5190, doi:10.1007/s00330-020-06787-9.
- [49] Schieda, N.; Nguyen, K.; Thornhill, R.E.; McInnes, M.D.F.; Wu, M.; James, N. Importance of Phase Enhancement for Machine Learning Classification of Solid Renal Masses Using Texture Analysis Features at Multi-Phasic CT. *Abdominal Radiology* 2020, 45, 2786–2796, doi:10.1007/s00261-020-02632-1.
- [50] Yap, F.Y.; Varghese, B.A.; Cen, S.Y.; Hwang, D.H.; Lei, X.; Desai, B.; Lau, C.; Yang, L.L.; Fullenkamp, A.J.; Hajian, S.; et al. Shape and Texture-Based Radiomics Signature on CT Effectively Discriminates Benign from Malignant Renal Masses. *Eur Radiol* 2020, 31, 1011–1021, doi:10.1007/s00330-020-07158-0.

A Hybrid Metaheuristic Model for Efficient Analytical Business Prediction

Marischa Elveny^{1*}, Mahyuddin K.M Nasution², Rahmad B.Y Syah³

Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia^{1,2}

Faculty of Engineering, Informatics Department, Universitas Medan Area, Medan, Indonesia³

Abstract—Accurate and efficient business analytical predictions are essential for decision making in today's competitive landscape. Involves using data analysis, statistical methods, and predictive modeling to extract insights and make decisions. Current trends focus on applying business analytics to predictions. Optimizing business analytics predictions involves increasing the accuracy and efficiency of predictive models used to forecast future trends, behavior, and outcomes in the business environment. By analyzing data and developing optimization strategies, businesses can improve their operations, reduce costs, and increase profits. The analytic business optimization method uses a hybrid PSO (Particle Swarm Optimization) and GSO (Gravitational Search Optimization) algorithm to increase the efficiency and effectiveness of the decision-making process in business. In this approach, the PSO algorithm is used to explore the search space and find the global best solution, while the GSO algorithm is used to refine the search around the global best solution. The hybrid meta-heuristic method optimizes the three components of business analytics: descriptive, predictive, and perspective. The hybrid model is designed to strike a balance between exploration and exploitation, ensuring effective search and convergence to high-quality solutions. The results show that the R2 value for each optimization parameter is close to one, indicating a more fit model. The RMSE value measures the average prediction error, with a lower error indicating that the model is performing well. MSE represents the mean of the squared difference between the predicted and optimized values. A lower error value indicates a higher level of accuracy.

Keywords—Efficiency; analytics business; predictions; Particle Swam Optimization (PSO); Gravitational Search Optimization (GSO)

I. INTRODUCTION

In the 4.0 Industrial Age, the availability of data is crucial for every strategic business decision [1-3]. Using analytics and algorithms, data is transformed into logical information [4]. In addition, data facilitates the consideration of visible and invisible problems in industrial operations [5-6]. Business analytics is the process of transforming data into valuable business knowledge using techniques and instruments [7-8]. Business Analytics accumulates historical business data, compiles, sorts, and then processes and analyzes the data using technology and company strategy in order to generate insights regarding company performance [9-11]. Business Analytics is a collection of techniques, technologies, and applications used to analyze company data and performance in order to make data-driven judgments regarding future investment strategies [12-13]. The three components of business analytics are descriptive analytics, the monitoring of

key performance indicators to understand current business conditions, predictive analytics, the analysis of trend data to predict possible future outcomes, and prescriptive analytics, the use of past performance to generate recommendations on how to handle similar situations in the future[14-17]. For optimization purposes in predictive business analytics, metaheuristics are applied.

The purpose of metaheuristics is to efficiently explore the search space in order to find the optimal solution. Metaheuristic techniques range from simple local search procedures to complex learning processes, from simple local search procedures to complex learning processes [18]. Gravitational Search Optimization (GSO) and Particle Swam Optimization (PSO) are both metaheuristic algorithms. Based on social behavior, PSO is an evolutionary algorithm. Populate the PSO algorithm's initial state with solutions [19]. The PSO algorithm incorporates a few performance-affecting parameters that are frequently expressed as an exploratory tradeoff [20]. Exploration is the ability to evaluate different regions of the problem space in pursuit of optimal solutions. PSO is frequently used to resolve multi-objective optimization issues [21-22]. This algorithm for solving complex problems has a simple yet effective strategy for optimizing numerical functions. GSO simulates interactions between objects in a search space, where objects represent candidate solutions and gravitational forces represent solution suitability to balance exploratory and exploitative search behavior [23-24].

II. RESEARCH METHODOLOGY

A. Methodology

The data is collected from e-metrics data. In pre-processing, the data is cleaned to remove inconsistencies and is null. Perform data integration and then transform the data to be ready for analysis. Feature selection and engineering for identification of variables impact business results, and engineering new features captures more information. It is followed by processing data using the PSO algorithm to optimize model performance using the GSO Algorithm. The training and test data are separated into training and test sets. Cross-validation was carried out for model robustness and to avoid overfitting. By implementing the model in a production environment, its performance is periodically monitored to detect changes in business conditions. The research steps can be seen in Fig. 1.

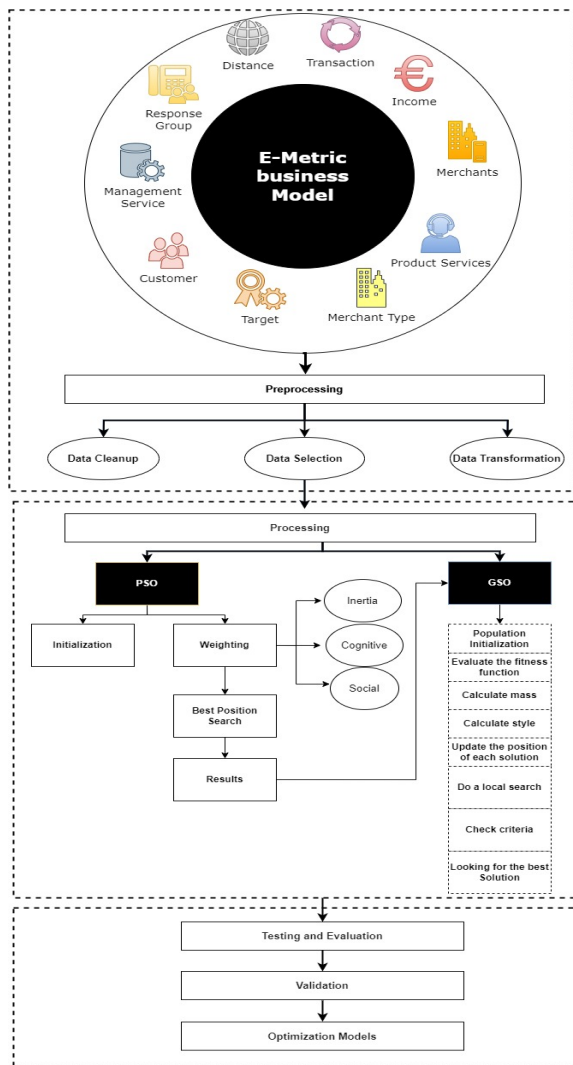


Fig. 1. Research methodology.

B. Definition of Predictive Decision Making

The process of making choices or decisions based on the analysis and interpretation of available data, patterns, and trends to forecast or predict future outcomes or events is known as predictive decision making [25]. Utilizing predictive analytics techniques and models to generate insights and estimates that can guide decision-making processes is included [26].

How can we develop accurate predictive models using machine learning techniques to forecast future business performance and identify potential optimization opportunities given the diversity of business datasets? The constructed model can account for a variety of business performance-influencing factors, such as market trends, customer behavior, and internal operations. Its purpose is to provide business stakeholders with insights and recommendations to assist them in making data-driven decisions and optimizing business operations [27]. Current trends emphasize the application of business analytics to forecasting. Fig. 2 depicts the solution to the problem, which is the optimized scope of the analytical business study.

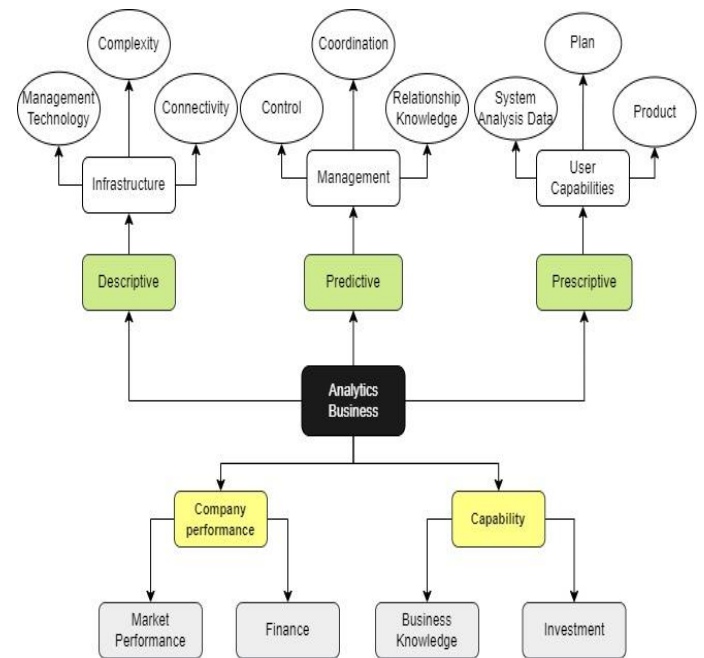


Fig. 2. Optimized analytical business review scope.

C. Classification of Data Analytics

Several advantages of big data analytics exist for obtaining valuable business insights. Here are some important benefits [28-29]:

- **Enhanced Choice Making:** This result is more precise and enlightened strategic planning, operational optimizations, and efficient resource allocation.
- **Competitive Advantage:** This information facilitates the identification of market opportunities, the development of effective marketing strategies, and the maintenance of a competitive advantage.
- **Enhanced Consumer Satisfaction:** This allows for targeted marketing, personalized recommendations, and superior customer experience, which ultimately increases customer satisfaction and loyalty.
- **Improved Risk Management:** This enables businesses to prevent fraud and mitigate risks, thereby safeguarding their assets and reputation.
- **Increased Productivity:** This contributes to the delivery of products and services that are more in line with consumer demands, thereby enhancing competitiveness and customer satisfaction.

The visualization is shown in Fig. 3.

D. Problem Solving Approach

In the context of optimizing business analytic predictions with the Multi-Attribute Method (MAM), Particle Swarm Optimization (PSO) can be utilized as a metaheuristic algorithm to find optimal solutions in complex search spaces. PSO is a population-based optimization technique that discovers the optimal solution by imitating the social behavior of a particle swarm.

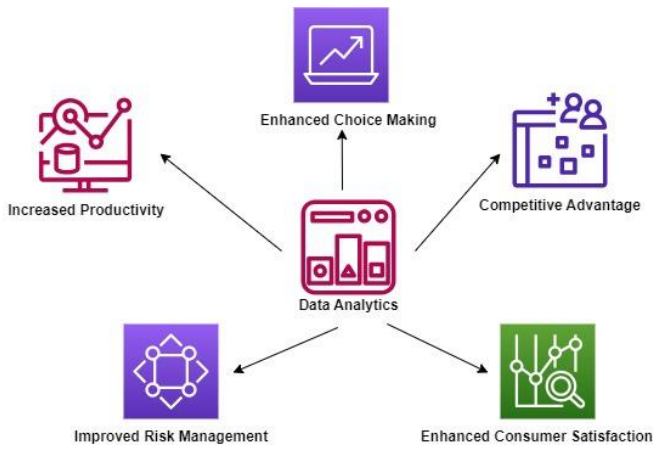


Fig. 3. Data analytics advantages for business insights.

1) Steps are done on Particle Swarm Optimization (PSO) [30]:

- Characteristics: Market conditions, customer behavior, financial indicators, and historical data are characteristics.
- Initialization of swarms: In the search space problem, the position and velocity of the particles are chosen at random.
- Fitness evaluation: Determine the fitness of each particle by calculating the value of the objective function based on the position of each particle. The fitness function indicates the efficiency or caliber of the particle solution.
- Update particle best position: Update each particle's position based on its current fitness.
- Update the highest global position: Determine the global best position by selecting the particle with the best position in the swarm.
- Update particle speed and position: Update the particle's speed and position.
- Repeat the process of fitness evaluation, updating particle best position, global best position, as well as velocity and position, until termination is reached.
- After the iteration is complete, extract the optimal solution considering the best position discovered.

To determine the optimal solution, the Particle Swarm Optimization (PSO) algorithm modifies the velocity and position of particles in the search space. Here are the formulas used by the conventional PSO algorithm [31].

PSO in training multilayer ceptrons' efficacy,

$$M_i = \{M_i^{[1]}, M_i^{[2]}\} (1)$$

Position indicates the optimal fitness value for every particle,

$$S_i = \{S_i^{[1]}, S_i^{[2]}\} (2)$$

seeking the optimal particle index at x ,

$$A_x = \{S_x^{[1]}, S_x^{[2]}\} (3)$$

Velocity update formula,

$$vl_{i(t+1)} = w * vl_{i(t)} + c_1 * rand() * (pbest_i - a_{i(t)}) + c_2 * rand() * (gbest - a_{i(t)}) (4)$$

Where,

- $vl_{i(t+1)}$ at time $t + 1$, is the updated velocity of particle i .
- w is the mass of inertia, which controls the influence of the previous velocity on the current velocity.
- $vl_{i(t)}$ at time t , the current velocity of particle i
- c_1 and c_2 acceleration coefficients that govern the effect of personal best ($pbest_i$) and global best ($gbest$) positions on the updated velocity.
- $rand()$ an arbitrary number between zero and one.
- $pbest_i$ is particle i 's personal best position, indicating the highest position it has attained to date.
- $a_{i(t)}$ at time t , is the present position of particle i .
- $gbest$ is the optimal position for all particles in the swarm globally.

Position update formula,

$$a_{i(t+1)} = a_{i(t)} + vl_{i(t+1)} (5)$$

In this formula, $a_{i(t+1)}$ represents particle i 's position updated at time $t + 1$, and $vl_{i(t+1)}$ the updated velocity calculated in the previous step.

In the Particle Swam Optimization (PSO) algorithm each particle moves towards its previous personal best position ($pbest$) and the global best position ($gbest$) to achieve the optimal solution, according to Eq. (6) [32].

$$p_{best_x}^i = B_x^* | f(B_x^*) = \min_{s=1,2,\dots,i} (\{B_x^s\}) (6)$$

where $x \in \{1, 2, \dots, N\}$

$$g_{best_x}^i = B_*^i | f(B_*^i) = \min_{s=1,2,\dots,i} (\{B_x^s\})$$

x represents the particle index, i represents the current iteration number, f represents the objective function to be optimized, g represents the position vector, and N represents the total number of particles within the flock. At each $x + 1$ iteration, the velocity D and position t of each particle i in the system are calculated. Eq. (7).

$$D_i^{x+1} = \omega D_i^x + v_1 r_1 (p_{best_i}^x - t_i^x) + v_2 r_2 (g_{best_i}^x - t_i^x) \\ t_i^{x+1} = t_i^x + D_i^{x+1} (7)$$

D represents the velocity vector, is used to balance local exploitation and global exploration, and v_1 and r_1 are uniformly distributed random vectors in the interval $[0,1]$. D are the dimensions of the search space or the magnitude of the encountered problem, and v_1 and v_2 are referred to as

"acceleration coefficients."

2) *Gravitational search optimization (GSO)*: The Gravitational Search Optimization (GSO) algorithm replicates the interaction between objects in the search space, where each object represents a candidate solution and gravitational force represents the solution's suitability [33].

The formula for updating the position of the *i*-th solution in the population of *N* solutions in the GSO algorithm in Eq. (8) and (9) [34].

$$a_i = G * \left(\frac{1}{d_i^2}\right) * (x_{cm} - x_i) \quad (8)$$

$$b_{i(t+1)} = b_{i(t)} + c_{i(t+1)} \quad (9)$$

Where,

- a_i Solution acceleration to- *i*, which is determined by the force of gravity acting on the solution.
- G The gravitational constant, which controls the strength of the gravitational force.
- d_i Euclidean distance between solutions to- *i* and the center of mass of the solution x_{cm}
- $b_{i(t)}$ Solution position to- *i* at time *t*
- $c_{i(t+1)}$ Updated speed from solution to- *i* at time *t* + 1
- $b_{i(t+1)}$ The latest position of the solution to- *i* at time *t* + 1

The approach to problem-solving that combines PSO and GSO is the PSO algorithm updates the position and velocity of the particle, then employs the GSO algorithm to update the fitness value and determine the optimal global position. The GSO algorithm can be used as an alternative to update the particle's position, while the PSO algorithm can be used to update the particle's velocity [35-37].

III. RESULT AND DISCUSSION

The first step is based on four segments. Segments are displayed based on the number of days that customers use to perform all activities from e-metric data. As in Table I, and the parameters used are market trends, behavior, customer needs, risk, service and product.

TABLE I. SEGMENT

Days	Segment
<200	1
200-500	2
500-1000	3
>1000	4

To successfully implement PSO, the optimal input parameter settings must be determined. The initial value and the final value govern the search process's exploration and exploitation. An explanation is provided in Table II.

TABLE II. DEFINITION OF THE PSO PROCESS

Definition	Information
Selected Data	x - direction, y - direction, z = 0.5, Q wall (77, 83, 110, 125)
Number of inputs in the best intelligence	5
Swarm Size (SS) in the best of intelligence (PSO parameter)	200
Changes in Accept Ratio (AR) which are evaluated (subtractive clustering parameter)	0.5, 0.6, 0.7, 0.8, 0.9
Changes in Inertia Weight Damping Ratio (WDR) which are evaluated(PSO parameter)	0.50, 0.60, 0.70, 0.80, 0.90
P(%) percentage of data which were used in training process(while 100% of data were considered in testing process	89%
Number of data	3546
Number of iterations	600

The objective value for finding the potential of each customer is calculated using Eq. (1). The results are shown in Table III.

TABLE III. POTENTIAL CUSTOMERS

Potential Customers				
	1	2	3	4
Market Trend	43.919	42.992	44.923	65.789
Behaviour	40.346	40.621	40.455	55.738
Customer Needs	54.748	53.637	43.637	56.738
Risk	53.737	57.728	53.637	63.789
Service	55.748	53.828	59.737	77.838
Product	50.763	53.728	60.738	79.748

Behavioral value is done by setting the customer's active power of each parameter based on the value of the objective function, best, worst and modifications to Eq. (2) and (3). i.e., the controlled variable related to customer status is modified to determine the best status in Eq. (6). The results are shown in Table IV.

TABLE IV. BEHAVIORAL VALUE

Values				
	1	2	3	4
Market Trend	0.536	-1536	0.637	0.647
Behaviour	0.787	-0.036	-1.748	0.537
Customer Needs	0.368	-0.546	0.074	0.647
Risk	-0.003	-0.002	0.004	0.003
Service	0.637	-1.738	0.647	0.663
Product	0.637	-1648	0.787	0.536

TABLE V. VELOCITY VALUE OPTIMIZATION

X[m]	Y[m]	Z[m]	Velocity [m s ⁻¹]	Velocity u[m s ⁻¹]	Velocity v[m s ⁻¹]	Velocity w[m s ⁻¹]	q wall
-0.003	-0.0040	0.30006	0.8130	-0.005	0.0002	0.81320	85
-0.008	-0.0383	0.30006	0.8680	-0.005	0.0002	0.86800	85
-0.0024	-0.0034	0.30006	0.8685	-0.005	0.0003	0.86875	85
-0.028	-0.0364	0.30006	0.8143	-0.004	0.0002	0.81413	85
-0.027	-0.0378	0.30006	0.7343	-0.004	0.0002	0.73483	85
-0.002	-0.0415	0.30006	0.7331	-0.004	0.0002	0.73381	85
-0.033	-0.0033	0.30006	0.7363	-0.003	0.0002	0.73623	85
-0.031	-0.0318	0.30006	0.8154	-0.004	0.0003	0.81534	85

For optimization, vectors are used as the particle representation. The population reacts to factors based on the highest individual and group scores. The distribution of responses between individual and group values ensures response diversity. The PSO algorithm instructs multi-layered perception in which the matrix learning problem is addressed. Eq. (4) to (7) is used to calculate the new velocity of a particle based on the particle's previous velocity and the distance from its current position, using the individual's and group's greatest experience. They demonstrate cooperation between particles within a collective. Then define a new position based on the new velocity as listed in Table V.

After the PSO results are obtained, it is optimized again using GSO. According to the GSO algorithm, gravitational and inertial masses are equivalent. However, the value employed is unique. When conducting search operations, the inertial mass increases because the movement becomes slower. As shown in Fig. 4 and Fig. 5, a greater gravitational mass causes a stronger attraction, allowing for a quicker convergence.

Multi-attribute evaluations provide comparable initial statuses, and the perception training procedure starts with the most suitable initial population, as shown in Table V. Performance is shown according to the age range of customers. Because age has an important effect on the behavior of each segment attribute. Age is divided into three parts, young age range 25-35, middle age 35-55 and old age 55-65 as shown in Table VI.

The eligibility of each customer is divided by age in finding each habit, in Eq. (8) and (9). Every solution must meet quality constraints as shown in Table VII behavior at a young age, Table VIII behavior at middle age and Table IX behavior at old age. It appears that the average middle age does more activity than the young and the old.

The augmentation of the hybrid method by incorporating a memory strategy with each individual's finest fitness history. In addition, the sensitivity analysis of the parameters in this instance is optimized according to the age of the customer. Finally, comparative experiments on a set of benchmark functions were conducted to assess the performance of the hybrid model. The results are presented in Table X.

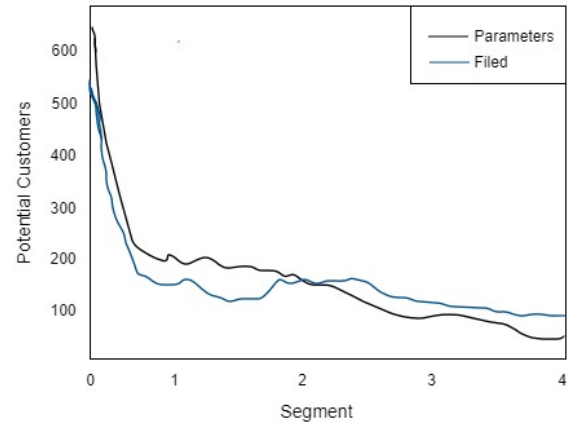


Fig. 4. Potential customer patterns.

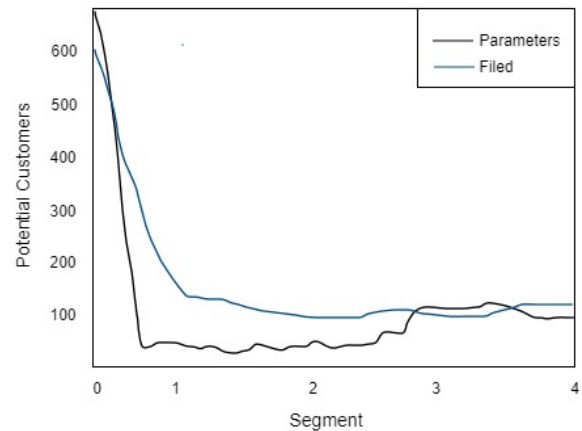


Fig. 5. Pattern of potential customers based on convergence.

TABLE VI. AGE CRITERIA

	Segments	Age Criteria
Young Age	1	25-35
Mid Age	2	35-55
Old Age	3	55-65

TABLE VII. YOUNG AGED CUSTOMERS

Young Aged Customers		
	Average Revenue	Expected Revenue
Low Income	<5,000	5,373,930
Medium Income	10,000	3,739,399
High Income	50,000	8,738,399
	Average	55,950,576

TABLE VIII. MID AGED CUSTOMERS

Mid Aged Customers		
	Average Revenue	Expected Revenue
Low Income	16,737	52,843,301
Medium Income	26,379	8,3286,416
High Income	37,585	118,667,120
	Average	849,324,224

TABLE IX. OLD AGED CUSTOMERS

Old Aged Customers		
	Average Revenue	Expected Revenue
Low Income	8,367	264,171,291
Medium Income	16,563	522,943,599
High Income	27,389	86,475,2897
	Average	506,225,957

TABLE X. HYBRID METHOD AUGMENTATION RESULTS

Customer Segments	Count of Purchase	Sum of Segmented Customers	Conversion Rate
1	27,153	27,153	33.74
2	36,538	36,849	47.62
3	83,638	25,379	44.63
4	73,647	53,838	34.58

IV. VALIDATION

Validation using R2, RMSE and MSE where R2 indicates the proportion of variance in the dependent variable that can be attributed to the independent variable in the regression model. It ranges from 0 to 1, with greater values indicating superior models [38]. RMSE is a measure of the average prediction error of a regression model. It is the square root of the average of the squared differences between predicted and observed values [39]. MSE is similar to RMSE, but it excludes the square root. It is the average of the squared deviations between predicted and actual values [40].

The formula to calculate R2 is [41]:

$$R2 = 1 - \left(\frac{SSR}{SST}\right) \quad (10)$$

Where:

- Sum of Squares Residual (SSR) is the sum of the squared differences between anticipated and observed values.

- SST (Total Sum of Squares) is the sum of the squared differences between actual values and the mean of the dependent variable.

The formula to calculate RMSE is [42]:

$$RMSE = \sqrt{\frac{1}{n} * \sum (y_i - y_{hat})^2} \quad (11)$$

Where:

- n the quantity of observations.
- y_i represents the exact value of the dependent variable.
- y_{hat} represents the value predicted for the dependent variable.

The formula to calculate MSE is [43]:

$$MSE = \frac{1}{n} * \sum (y_i - y_{hat})^2 \quad (12)$$

Where:

- n is the number of occurrences.
- y_i represents the exact value of the dependent variable.
- y_{hat} represents the value predicted for the dependent variable.

Table XI displays the results, where the R2 value for each segment is close to one, indicating a superior model fit. The RMSE value assesses the average error in prediction, with a lower value indicating that the model is performing well. MSE is the mean of the squared differences between predicted and optimized values. A lower value indicates a higher level of precision.

TABLE XI. R2, RMSE, MSE VALIDATION RESULTS

Parameter		Training			Validation		
Customer Segments	Conversion Rate	R2	RMSE	MSE	R2	RMSE	MSE
1	33,74	0.87	0.17	0.08	0.87	0.20	0.11
2	47,62	0.97	0.26	0.29	0.93	0.27	0.21
3	44,63	0.85	0.36	0.28	0.91	0.41	0.35
4	34,58	0.89	0.45	0.22	0.89	0.40	0.43

Fig. 6 and Fig. 7 illustrate that the PSO Algorithm's inputs and parameters are based on social behavior, whereas the GSO is based on mass physical phenomena. In PSO, each particle modifies its position based on its own optimal position and the optimal position of the entire system. GSO, each agent's position changes dependent on the combined power of all other agents. PSO utilizes memory to update the velocity and position of particles. The GSO acceleration of the agent has an effect on the position and velocity updates. PSO particles' positions are updated without regard to the distance between solutions, whereas GSO particles' positions are updated using a force that is inversely proportional to the distance between solutions. The obtained results demonstrate that PSO improves every customer's social behavior based on their needs, current trends, risks, and services whereas GSO optimizes every condition consideration for future improvement.

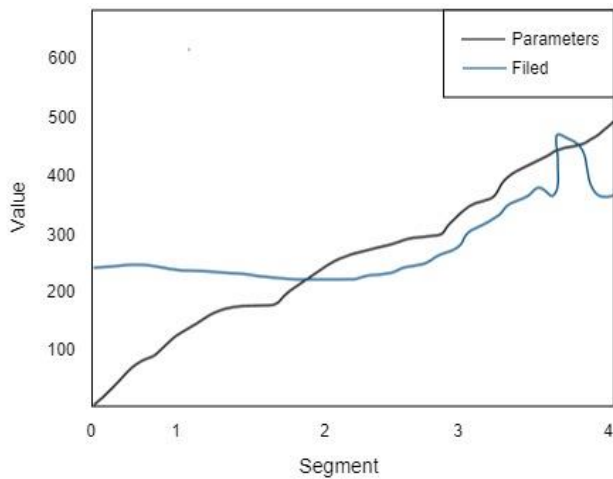


Fig. 6. Inputs and parameters of PSO and GSO algorithms are based on social behaviors.

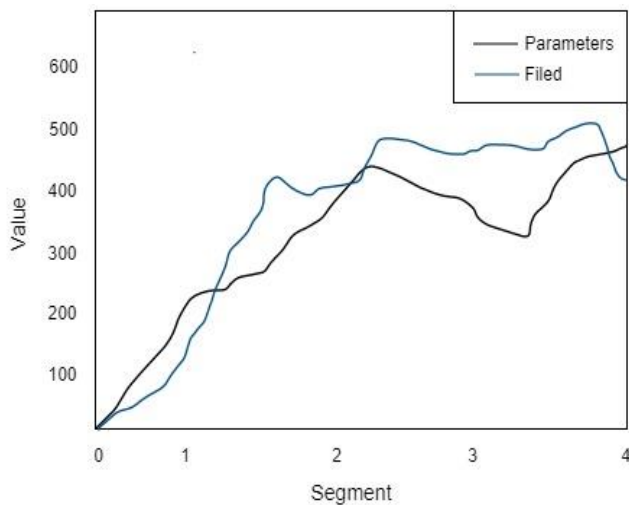


Fig. 7. The results of the optimization of the PSO and GSO algorithms based on the value social behaviors.

V. CONCLUSION

The conclusion of this study is that PSO and GSO are efficient meta-heuristic optimization algorithms for business analysis by augmenting a specific decision-making process or business outcome. In terms of convergence speed, precision, robustness, or scalability, the performance of PSO and GSO is analyzed. It is evidenced by the R2 validation value close to one, RMSE and MSE with lower error rates. Thus, increasing the effectiveness of these business analytics can overcome any limitations or barriers associated with the practical application of algorithms in a business environment.

ACKNOWLEDGMENT

Thanks to DRTPM for the research funding provided with contract no. 70/UN5.2.3.1/PPM/KP-DRTPM/B/2023.

REFERENCES

[1] R. Syah, M. Elveny, and M. K. M. Nasution, "Clustering Large DataSet to Prediction Business Metrics," 2020, pp. 1117–1127. doi: 10.1007/978-3-030-63322-6_95.

[2] R. Syah, M. Elveny, M. K. M. Nasution, and G. W. Weber, "Enhanced Knowledge Acceleration Estimator Optimally with MARS to Business Metrics in Merchant Ecosystem," in 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/ELTICOM50775.2020.9230487.

[3] L. S. Dalenogare, G. B. Benitez, N. F. Ayala, and A. G. Frank, "The expected contribution of Industry 4.0 technologies for industrial performance," *Int J Prod Econ*, vol. 204, pp. 383–394, Oct. 2018, doi: 10.1016/j.ijpe.2018.08.019.

[4] H.-N. Dai, H. Wang, G. Xu, J. Wan, and M. Imran, "Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies," *Enterp Inf Syst*, vol. 14, no. 9–10, pp. 1279–1303, Nov. 2020, doi: 10.1080/17517575.2019.1633689.

[5] R. H. Hamilton and W. A. Sodeman, "The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources," *Bus Horiz*, vol. 63, no. 1, pp. 85–95, Jan. 2020, doi: 10.1016/j.bushor.2019.10.001.

[6] N. A. Ghani, S. Hamid, I. A. Targio Hashem, and E. Ahmed, "Social media big data analytics: A survey," *Comput Human Behav*, vol. 101, pp. 417–428, Dec. 2019, doi: 10.1016/j.chb.2018.08.039.

[7] C. Yang et al., "Big Earth data analytics: a survey," *Big Earth Data*, vol. 3, no. 2, pp. 83–107, Apr. 2019, doi: 10.1080/20964471.2019.1611175.

[8] M. Kraus, S. Feuerriegel, and A. Oztekin, "Deep learning in business analytics and operations research: Models, applications and managerial implications," *Eur J Oper Res*, vol. 281, no. 3, pp. 628–641, Mar. 2020, doi: 10.1016/j.ejor.2019.09.018.

[9] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," *Comput Chem Eng*, vol. 126, pp. 465–473, Jul. 2019, doi: 10.1016/j.compchemeng.2019.04.003.

[10] S. Ren, Y. Zhang, Y. Liu, T. Sakao, D. Huisingh, and C. M. V. B. Almeida, "A comprehensive review of big data analytics throughout product lifecycle to support sustainable smart manufacturing: A framework, challenges and future research directions," *J Clean Prod*, vol. 210, pp. 1343–1365, Feb. 2019, doi: 10.1016/j.jclepro.2018.11.025.

[11] B. Bajic, A. Rikalovic, N. Suzic, and V. Piuri, "Industry 4.0 Implementation Challenges and Opportunities: A Managerial Perspective," *IEEE Syst J*, vol. 15, no. 1, pp. 546–559, Mar. 2021, doi: 10.1109/JSYST.2020.3023041.

[12] M.; N. M. K. M.; Z. M.; E. S. Elveny, "An advantage optimization for profiling business metrics competitive with robust nonparametric regression," *J Theor Appl Inf Technol*, vol. 99, pp. 114–124, 2021.

[13] M. K. M. Nasution, O. S. Sitompul, M. Elveny, and R. Syah, "Data science: A Review towards the Big Data Problems," *J Phys Conf Ser*, vol. 1898, no. 1, p. 012006, Jun. 2021, doi: 10.1088/1742-6596/1898/1/012006.

[14] A. Al-Khowarizmi, R. Syah, and M. Elveny, "The Model of Business Intelligence Development by Applying Cooperative Society Based Financial Technology," 2022, pp. 117–125. doi: 10.1007/978-981-16-2377-6_13.

[15] R. Syah, M. Elveny, and M. K. M. Nasution, "Performance Knowledge Acceleration Optimization with MARS to Customer Behavior in Merchant Ecosystem," in 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), IEEE, Jun. 2020, pp. 178–182. doi: 10.1109/MECnIT48290.2020.9166604.

[16] Z. Huang, K. S. Savita, and J. Zhong-jie, "The Business Intelligence impact on the financial performance of start-ups," *Inf Process Manag*, vol. 59, no. 1, p. 102761, Jan. 2022, doi: 10.1016/j.ipm.2021.102761.

[17] M.-L. Tseng, T. P. T. Tran, H. M. Ha, T.-D. Bui, and M. K. Lim, "Sustainable industrial and operation engineering trends and challenges Toward Industry 4.0: a data driven analysis," *Journal of Industrial and Production Engineering*, vol. 38, no. 8, pp. 581–598, Nov. 2021, doi: 10.1080/21681015.2021.1950227.

[18] S. Mahdinia, M. Rezaie, M. Elveny, N. Ghadimi, and N. Razmjooy, "Optimization of PEMFC Model Parameters Using Meta-Heuristics," *Sustainability*, vol. 13, no. 22, p. 12771, Nov. 2021, doi: 10.3390/su132212771.

- [19] S. S. Band et al., "Novel Ensemble Approach of Deep Learning Neural Network (DLNN) Model and Particle Swarm Optimization (PSO) Algorithm for Prediction of Gully Erosion Susceptibility," *Sensors*, vol. 20, no. 19, p. 5609, Sep. 2020, doi: 10.3390/s20195609.
- [20] E. Bas, E. Egrioglu, and E. KOLEMEN, "Training simple recurrent deep artificial neural network for forecasting using particle swarm optimization," *Granular Computing*, vol. 7, no. 2, pp. 411–420, Apr. 2022, doi: 10.1007/s41066-021-00274-2.
- [21] M. Jain, V. Saijpal, N. Singh, and S. B. Singh, "An Overview of Variants and Advancements of PSO Algorithm," *Applied Sciences*, vol. 12, no. 17, p. 8392, Aug. 2022, doi: 10.3390/app12178392.
- [22] T. M. Shami, A. A. El-Saleh, M. Alswaiti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, "Particle Swarm Optimization: A Comprehensive Survey," *IEEE Access*, vol. 10, pp. 10031–10061, 2022, doi: 10.1109/ACCESS.2022.3142859.
- [23] M. S. Nazir et al., "Optimized economic operation of energy storage integration using improved gravitational search algorithm and dual stage optimization," *J Energy Storage*, vol. 50, p. 104591, Jun. 2022, doi: 10.1016/j.est.2022.104591.
- [24] R. Shankar, N. Ganesh, R. Čep, R. C. Narayanan, S. Pal, and K. Kalita, "Hybridized Particle Swarm—Gravitational Search Algorithm for Process Optimization," *Processes*, vol. 10, no. 3, p. 616, Mar. 2022, doi: 10.3390/pr10030616.
- [25] S. Arena, E. Florian, I. Zennaro, P. F. Orrù, and F. Sgarbossa, "A novel decision support system for managing predictive maintenance strategies based on machine learning approaches," *Saf Sci*, vol. 146, p. 105529, Feb. 2022, doi: 10.1016/j.ssci.2021.105529.
- [26] M. Achouch et al., "On Predictive Maintenance in Industry 4.0: Overview, Models, and Challenges," *Applied Sciences*, vol. 12, no. 16, p. 8081, Aug. 2022, doi: 10.3390/app12168081.
- [27] Q. Cao et al., "KSPMI: A Knowledge-based System for Predictive Maintenance in Industry 4.0," *Robot Comput Integr Manuf*, vol. 74, p. 102281, Apr. 2022, doi: 10.1016/j.rcim.2021.102281.
- [28] H. Zhang, Z. Zang, H. Zhu, M. I. Uddin, and M. A. Amin, "Big data-assisted social media analytics for business model for business decision making system competitive analysis," *Inf Process Manag*, vol. 59, no. 1, p. 102762, Jan. 2022, doi: 10.1016/j.ipm.2021.102762.
- [29] J. Yang, P. Xiu, L. Sun, L. Ying, and B. Muthu, "Social media data analytics for business decision making system to competitive analysis," *Inf Process Manag*, vol. 59, no. 1, p. 102751, Jan. 2022, doi: 10.1016/j.ipm.2021.102751.
- [30] N. Y. Vanguri, S. Pazhanirajan, and T. A. Kumar, "Competitive feedback particle swarm optimization enabled deep recurrent neural network with technical indicators for forecasting stock trends," *Int J Intell Robot Appl*, vol. 7, no. 2, pp. 385–405, Jun. 2023, doi: 10.1007/s41315-022-00250-2.
- [31] Y. Luo, J. Zhong, W.-L. Liu, and W.-N. Chen, "Automatic Business Location Selection through Particle Swarm Optimization and Neural Network," in *2023 15th International Conference on Advanced Computational Intelligence (ICACI)*, IEEE, May 2023, pp. 1–8. doi: 10.1109/ICACI58115.2023.10146157.
- [32] X. Zhang, H. Liu, and L. Tu, "A modified particle swarm optimization for multimodal multi-objective optimization," *Eng Appl Artif Intell*, vol. 95, p. 103905, Oct. 2020, doi: 10.1016/j.engappai.2020.103905.
- [33] S. K. , & P. D. R. Mohapatra, "A new gravitational search algorithm for global optimization," *Neural Comput Appl*, vol. 3, no. 29, pp. 713–733, 2018.
- [34] C. Subbalakshmi, P. K. Pareek, and M. V. Narayana, "A Gravitational Search Algorithm Study on Text Summarization Using NLP," 2022, pp. 144–159. doi: 10.1007/978-3-031-21385-4_13.
- [35] A. Singh and N. Singh, "Gravitational search algorithm-driven missing links prediction in social networks," *Concurr Comput*, vol. 34, no. 11, May 2022, doi: 10.1002/cpe.6901.
- [36] N. Ayyash and F. Hejazi, "Development of hybrid optimization algorithm for structures furnished with seismic damper devices using the particle swarm optimization method and gravitational search algorithm," *Earthquake Engineering and Engineering Vibration*, vol. 21, no. 2, pp. 455–474, Apr. 2022, doi: 10.1007/s11803-022-2088-1.
- [37] X. Hou, S. Gao, L. Qiu, Z. Li, R. Zhu, and S.-K. Lyu, "Transmission Efficiency Optimal Design of Spiral Bevel Gear Based on Hybrid PSO-GSA (Particle Swarm Optimization—Gravitational Search Algorithm) Method," *Applied Sciences*, vol. 12, no. 19, p. 10140, Oct. 2022, doi: 10.3390/app121910140.
- [38] A. Gomez-Flores, S. A. Bradford, L. Cai, M. Urík, and H. Kim, "Prediction of attachment efficiency using machine learning on a comprehensive database and its validation," *Water Res*, vol. 229, p. 119429, Feb. 2023, doi: 10.1016/j.watres.2022.119429.
- [39] M. H. Rashidi, S. Keshavarz, P. Pazarí, N. Safahieh, and A. Samimi, "Modeling the accuracy of traffic crash prediction models," *IATSS Research*, vol. 46, no. 3, pp. 345–352, Oct. 2022, doi: 10.1016/j.iatssr.2022.03.004.
- [40] M. R. Machado and S. Karray, "Assessing credit risk of commercial customers using hybrid machine learning algorithms," *Expert Syst Appl*, vol. 200, p. 116889, Aug. 2022, doi: 10.1016/j.eswa.2022.116889.
- [41] Z. Huang, K. S. Savita, and J. Zhong-jie, "The Business Intelligence impact on the financial performance of start-ups," *Inf Process Manag*, vol. 59, no. 1, p. 102761, Jan. 2022, doi: 10.1016/j.ipm.2021.102761.
- [42] U. Gupta, V. Bhattacharjee, and P. S. Bishnu, "StockNet—GRU based stock index prediction," *Expert Syst Appl*, vol. 207, p. 117986, Nov. 2022, doi: 10.1016/j.eswa.2022.117986.
- [43] Z. Fathali, Z. Kodia, and L. Ben Said, "Stock Market Prediction of NIFTY 50 Index Applying Machine Learning Techniques," *Applied Artificial Intelligence*, vol. 36, no. 1, Dec. 2022, doi: 10.1080/08839514.2022.2111134

A Mechanism for Bitcoin Price Forecasting using Deep Learning

Karamath Ateeq¹, Ahmed Abdelrahim Al Zarooni², Abdur Rehman³, Muhammd Adna Khan⁴

School of Computing, Skyline University College, Sharjah, United Arab Emirates^{1,2,4}

Riphah School of Computing & Innovation-Faculty of Computing, Riphah International University, Lahore, Pakistan^{3,4}

Department of Software-Faculty of Artificial Intelligence and Software, Gachon University, Seongnam-si, Republic of Korea⁴

Abstract—Researchers and investors have recently become interested in forecasting the cryptocurrency price forecasting but the most important currency can take that it's the bitcoin exchange rate. Some researchers have aimed at leveraging the technical and financial characteristics of Bitcoin to create predictive models, while others have utilized conventional statistical methods to explain these factors. This article explores the LSTM model for forecasting the value of bitcoins using historical bitcoin price series. Predict future bitcoin prices by developing the most accurate LSTM forecasting model, building an advanced LSTM forecasting model (LSTM-BTC), and comparing past bitcoin prices. This is the second step, if looking at the end of the model, it has very high accuracy in predicting future prices. The performance of the proposed model is evaluated using five different datasets with monthly, weekly, daily, hourly, and minute-by-minute bitcoin price data with total records from January 1, 2021, to March 31, 2022. The results confirm the better forecasting accuracy of the proposed model using LSTM-BTC. The analysis includes square error MSE, RMSE, MAPE, and MAE of bitcoin price forecasting. Compared to the conventional LSTM model, the suggested LSTM-BTC model performs better. The contribution made by this research is to present a new framework for predicting the price of Bitcoin that solves the issue of choosing and evaluating input variables in LSTM without making firm data assumptions. The outcomes demonstrate its potential use in applications for industry forecasting, including different cryptocurrencies, health data, and economic time.

Keywords—Currency; bitcoin; LSTM; forecasting; models

I. INTRODUCTION

Since cryptocurrencies are the newest financial innovation and are having a significant impact on the world economy, cryptocurrency price forecasts are very important. Fintech professionals and technologists are particularly interested in forecasting the price of cryptocurrencies and hosting blockchain conferences to educate the public on the most recent revolution. Previous studies have observed proof of this link between modifications in stock prices and social media [1]. Cryptocurrency is a digital trade concept that makes use of cryptographic capabilities to conduct economic transactions. Cryptocurrencies leverage the blockchain era to the advent of age transparency, decentralization, and immutability. BTC is the maximum well-known cryptocurrency, which came into existence in 2009 through an anonymous institution or person, accomplishing its height cost on December 16, 2017, through mountaineering to nearly \$20,000. In the final ten years, 1512 alternative cryptocurrencies like Ethereum and Litecoin had

been created proving that the cryptocurrency marketplace has emerged revealing its sturdy growth [2]. The forecasting of Bitcoin price may be taken into consideration as a common sort of time-limited problem, just like the stock price prediction. Traditional time-series models, just like the famous ARIMA, had been carried out for cryptocurrencies' price and motion prediction [4,6]. After the discovery of the Blockchain era, which began approximately a decade ago, a maximum of the posted studies on this region have been targeting non-technological factors of Blockchain technology inclusive of felony troubles and its function in criminal activities [7]. Since cryptocurrencies have been first added in the year 2008 their scope has been restricted to papers posted between 2008 and 2018 [8]. This paper discusses the research methodologies used in this study and displays the result lot of the valuation of reviewed papers and their classifications. RNNs that upload the particular dealing with the sequence of observations include the LSTM. Whilst learning a mapping feature from inputs to outputs, is now no longer supplied through MLPs or CNNs [9-10]. Most commonly, statistical strategies are being in use for such a long time, from the 1970s onward, exclusively the ones primarily based totally on one Box-Jenkins methodology [11]. By reviewing rising studies of deep-learning fashions, which include their mathematical formulation, for large facts, and function learning. Another terrific work may be determined wherein the authors added the time series type hassle and furnished an open-supply framework with applied algorithms in the University of East Anglia/the University of California in Riverside source [12]. The forecasting trouble and mathematical method for time collection may be determined in the Problem Definition segment [13]. The deep-Learning Architecture phase presents the deep-studying architectures normally used in the context of time collection forecasting. A time collection is described as a chain of values determined over time. As it is a known term that time is a variable measured on a non-stop basis, the values in a time collection are tested at consistent intervals (constant sampling frequency) [14,15].

Cryptocurrencies have attracted recognition as volatile investments, because of excessive investor losses due to scams, hacks, and bugs. Although the underlying cryptography is usually secure, the technical complexity of the use of and storing crypto belongings may be a primary threat to new users. Unlike conventional finance, there may be no manner to do the opposite or cancel a cryptocurrency transaction after it has already been sent. The statistics on the blockchain are encrypted, which means no one can mess with it. During

transactions the person's Sample paragraph, the complete file ought to be in the name isn't revealed, however simplest their wallet ID is made public. At present, the costs of those cryptocurrencies do now no longer have a good-sized quantity of research and studies in comparison to standard trading markets. The process of Forecasting Time Series has continually become a crucial region of research in lots of domain names due to the fact many specific forms of information are saved as time collection [16]. For example, by discovering quite a few times collection information in medication, climate forecasting, supply chain control, biology and forecasting of stock prices, etc. with knowledge resources of developing availability of information and computing strength in the current years, Deep Learning has ended up an essential a part of the new technology of Time Series Forecasting fashions, acquiring tremendous results.

Deep learning can be the future of complicated and hard-time collection forecasting and the article will assist you to get commenced and make fast development for your personal forecasting problems. Supervised studying is wherein you have input variables (X) and an output variable (y), and you operate a set of rules to analyze the mapping characteristic from the entry to the output. Deep Learning used for forecasting Time Series overcomes the conventional Machine Learning risks with several special methods. In addition, don't forget the overall performance of the current attention-based Transformer models, which has had exact fulfillment withinside the image processing and herbal language processing domains. In all, by evaluating four special deep learning techniques (RNN, LSTM, GRU, and Transformer) at the side of a baseline approach. The process of implementing multilayer perceptron is enormously meek. However, deep-learning models are extra complex, along with their implementation calls for an excessive stage of technical information and substantial time investment to implement [17].

The authors of this research suggest a revolutionary machine-learning methodology for forecasting Bitcoin's price and behavior. A deep neural network underlies the proposed model, which was trained using a sizable dataset of historical Bitcoin price data. The model beats state-of-the-art techniques when tested on a held-out test set. Investors can utilize this model to aid them in making more educated choices regarding Bitcoin trading.

II. RELATED WORK

A cryptocurrency is a virtual foreign money, that is an opportunity shape of charge created by the usage of encryption algorithms. The use of encryption technology methods that cryptocurrencies characteristic each as foreign money and as a digital accounting system. To use cryptocurrencies, you want a cryptocurrency wallet. These wallets may be the software program that may be a cloud-based provider or is saved in your computer or cellular device. The wallets are the device via which you store your encryption keys that verify your identification and link to your cryptocurrency. In these wallets, there are two types of keys. The first one is a public key that is visible to all and helps you to identify the transaction being made from your account, or in simple words, represents your account publicly. While the other the most sensitive key is

known as the private key which in term helps you to send and receive transactions inside the blockchain. These wallets are in term only and can be used for trading in blockchain and more work as a bank account for your digital currency. What are the dangers of the usage of cryptocurrency? Cryptocurrencies are nonetheless fairly new, and the marketplace for those virtual currencies may be very volatile. Since cryptocurrencies do not want banks or some other third party to adjust them; they tend to be uninsured and are tough to transform right into a form of actual foreign money (which includes US greenbacks or euros.) In addition, considering that cryptocurrencies are technology-based intangible assets, they may be hacked like some other intangible generation assets. Finally, because you shop your cryptocurrencies in a virtual pocket, in case you lose your pockets (or access to it or pocket backups), you have misplaced your whole cryptocurrency investment.

Unlike conventional finance, there may be no manner to oppose or cancel a cryptocurrency transaction after it has already been sent. By a few estimates, approximately a 5th of all bitcoins is inaccessible because of misplaced passwords or wrong sending addresses. Bitcoin is a currency but not the basic one, it is a crypto forex that is used globally for virtual fees or truly used for investment. It is decentralized for example it isn't owned by anyone. Transactions made via way of means of Bitcoins are smooth as they may be observed in any country. Investment may be achieved thru numerous marketplaces recognized as "bitcoin exchanges".

A range of machine learning methods, such as sentiment analysis of Twitter feeds, have been developed over the years to forecast price movement for the financial markets using digital platforms. Recent research has effectively used sentiment analysis for a variety of purposes, including predicting movie box office receipts [1]. Machine learning, like deep learning techniques, is recognized as an efficient forecaster for several tasks and situations; such a toolkit provides algorithmic traders with a powerful yet fundamental set of tools to anticipate the path of price yet for capital assets. Huang, Xin, and others [2] described the purpose of LSTM Driven Sentiment Classification for Cryptocurrency Purpose of forecasting is to outline and illustrate the work of LSTM Results using data sets from social media posts, tweets, and comments. The concluded precision is 92.5 percent. In contrast to the conventional auto-regressive technique, the system used in this research is constructed via LSTM and attains better recall and precision. Wei Chen and others [3] used the data set which was gathered utilizing websites, APIs, and machine learning models. It was employed to forecast the Bitcoin exchange rate using technological and financial factors. Here, the data will be divided into 4 distinct periods concerning the year and the currency exchange rate using the methods ARIMA, SVR, ANFIS, and LSTM. Learning techniques that have been put into use according to research, the GRU model time series offers the fastest compilation of bitcoin price predictions [5].

T. Awoke et al. [10] said Bitcoin price prediction and analysis used a deep learning model where the dataset implements LSTM and GRU methods using Kaggle bitcoin price data (2014-2018 low price data) reaches up to 92% and 75% respectively. Long-term dependencies can be more

effectively identified using LSTM and GRU models. Some recent work has focused on high-frequency trading and the application of deep learning methods such as RNN to predict time series data whose functional models have been transformed into dense and feeder networks [18]. Deep learning methods are expected to outperform the deficient performance of ARIMA. McNally [19] used machine learning techniques such as recurrent neural networks (RNN) and long-term memory (LSTM) to predict the process of changing the price of Bitcoin and automatically compare the results. To enhance outcomes, such as deep learning employing neural networks (such as ANN and RNN) in prediction, apply ML algorithms (such as SVM, Bayesian Network, regress, or any other advanced machine learning approach) [20]. The same technique may be used by Hoy to forecast Bitcoin.

The market for algorithmic trading is estimated to be worth \$11.66 billion in 2020 and to increase to \$26.27 billion by 2021, according to research [21]. Bitcoin is the maximum famous instance of chain technology. "Bitcoin is a peer-to-peer digital coins system" delivered withinside the famous paper Nakamoto[22]. The peer-to-peer (P2P) mechanism permits a possession switch, from one party to every other without a third-party intervention (monetary institution). Payments may be remodeled on the internet with nonneutral or fee of a government for the primary time [23]. Deep Learning models examine capabilities and dynamics best and at once from the information. Thanks to this method, they accelerate the manner of information training and might examine extra complicated information styles in a greater whole way [24]. Before talking about Deep Learning techniques used for Forecasting Time Series, it's miles beneficial to keep in mind that the maximum classical Machine Learning models which are used to resolve this hassle are ARIMA models and additionally exponential smoothing [25]. The cryptocurrency marketplace is one of the quickest developing in the global and is taken into consideration as one of the maximum risky markets for transactions. For example, the price of bitcoins has skyrocketed, from nearly 0 in 2013 to around \$ 19,000 in 2017. For a few altcoins, the charge can upward thrust or fall through greater than 50% in a single day [26].

Deep learning models are one of, if now no longer the most data-hungry models of the Machine Learning world. They want big quantities of facts to attain their greatest overall performance and serve with the distinction, anticipate from them. However, having these many facts isn't always continually easy [27]. The purpose of machine learning is to discover styles in statistics and then make predictions, typically based on complicated styles, to reply to commercial enterprise questions, track trends, and assist them to examine and solve issues. Thus, it's far vital to look at it to discover an extensive variety of verbal facts more easily [28]. LSTM achieved overall recognition accuracy of 52% and 8% RMSE. Compared to deep learning systems, the popular ARIMA method is used for time series forecasting. This model is inefficient as deep learning models can be implemented.

III. LIMITATIONS OF PREVIOUS WORK AND OUR CONTRIBUTION

Past exchange rates are used as predictors in current research on bitcoin exchange rate forecasts. Bitcoin's creation, however, demonstrates a complicated and extremely erratic character. Therefore, it is crucial to take into account the many variables that affect the exchange rate for Bitcoin. A variety of econometric techniques, including random forests, clustering, and machine learning, have been used to investigate the factors that influence the price of Bitcoin. Here is Table I that shows us 'Previous Works' and the same method that is going to be utilized in this research for getting good results with the Forecast Data Set.

TABLE I. ANALYSIS OF PREVIOUS METHODS AND RESULTS WITH LSTM

Approach	Year	Dataset Period	Method used	Performance Evaluation
Huang et al. [2]	2021	19/03/2021 to 27/03/2021	LSTM	0.87
Fan Fang et al. [4]	2021	02/07/2018 to 29/08/2018	LSTM	0.82
Awoke et al. [10]	2020	01/01/2014-20/02/2018	LSTM	0.092
M.J. Hamayel et al. [11]	2021	01/01/2018 to 30/06/2021	LSTM	410.399

A. Our Contribution

The Long-Term Memory Network (LSTM) is a modern deep-learning architecture for time series estimates. Moreover, there hasn't been much research done on forecasting financial time series, particularly for cryptocurrencies. To predict the daily price of bitcoin using the usual LSTM model, offer a novel forecasting framework. The suggested model's effectiveness is assessed. By employing techniques relevant to the specific item, need to anticipate the value of Bitcoin. For instance, might simply wish to predict based on signals or prices, or might only be able to anticipate the current day based on LSTM, past prices, and other techniques (such as official forecasts). A closing price might also be anticipated the next day. Market trends might aid investors in choosing their investments. Be extremely precise when determining future values. Additionally, both in accuracy and precision, the LSTM model outperforms conventional LSTMs and other time series forecasting models. It uses sophisticated algorithms, statistical models, and human oversight to make trade choices on exchanges. In contrast, the Financial Times reports that, in addition to conventional hedge strategies and futures trading, quantitative trading is being utilized to trade digital currencies like Bitcoin. In-depth discussions are held about the development and assessment of statistical models, predicting, controlling, and filtering of ideal time series. The two main types of time series analysis are hyperbolic and random analysis of variance.

IV. MATERIALS AND METHODS

A. Research Methodology

The purpose of this research is to predict price changes in Crypto (Bitcoin) on the premise of Monthly, Weekly, Days, Hourly, and by Minutes time to alternate in price like multilevel (for example, the diploma of increase/decrease) and

binary (up / down). The transaction topic gives a Deep Learning Model's Techniques to predict the price dynamics of digital currencies in actual time by using the evaluation of numerous values. The LSTM cell has recall and overlooking gates, which allow the cell to choose which facts to propagate or block based on their relevance and strength [29]. Determine which information needs to be kept or forgotten, then upload it to the cell state or delete it. By enabling the recovery of data that has been copied into memory, LSTM can resolve the vanishing gradient problem [30]. Time series with unknown time delays may be classified, processed, and predicted with the help of LSTM.

B. Data Collection

Data preparation is the process of gathering, integrating, organizing, and structuring records so that they may be used for applications such as data analytics, data mining, and visualization. For the issue we're trying to address, accurate records must be fed. The preparation of the data set is a crucial stage in machine learning. As previously said, the data preparation has an impact on the accuracy of the forecasts; as a result, this part has to explain the information in the records set. The methods employed to get the data ready for the model being utilized. For the main part of this research, the publicly accessible API from BINANCE Exchange was used to gather historical and real-time rate data for Bitcoin. By collecting real-time bitcoin data from the Coinbase API, allowing one to predict price fluctuations concerning other datasets.

C. Pre-Processing

The dataset was split into training and test subsets at a ratio of 90:10 to train the models, fit them, and fine-tune their parameters. This dataset was built up over each prospect's period and continued the data-collecting process. The facts Set is existed on 5 exclusive forms and LSTM BTC Model is implemented on those all. All of the facts set are from January 2021 to March 2022. The First Data Set is set Monthly Basis Data set. The Second is set to Weekly, 0.33 is set everyday basis, the fourth is about Hourly, and the Last one is an approximately Minutes-by-minutes data set. By anticipating the Bitcoin price for the next time using the Minutes inside hours, weekly, monthly, daily, and hourly prices as input. To improve statistics processing and version convergence efficiency, a variety of pre-processing approaches are used. Minibatch is used to divide massive statistics into manageable batches, which boosts memory effectiveness. Some data in the statistics utilized in this study were missing or had no relevance because they were collected through websites and APIs.

D. Proposed Model

It is claimed that the vanishing gradient problem is avoided by a deep learning concept, particularly a Recurrent Neural Network notion. The main reason for using this set of rules is that it prevents back propagation errors from disappearing or bursting, allowing them to travel backward via an endless number of digital layers that have been opened up in space. The research is divided into three sections due to the heritage records and restricted circumstances. Information preprocessing comes first. Use a modified LSTM network for forecasting and training. The use of interpolation fitting and Fourier rework noise discount for Bitcoin price data, to

improve accuracy in the later time collection prediction. To improve forecasting results, cellular connections are added to the candidate hidden states and control gate in the unique LSTM version, and the most effective control gate is maintained. Following the procedure of purging the data set and dividing it into train data sets, LSTM-BTC is used to observe the results for each of this research paper's data sets. The specific flowchart for the entire essay can be seen in Fig. 1 to avoid using complicated descriptions and to intuitively duplicate this work method.

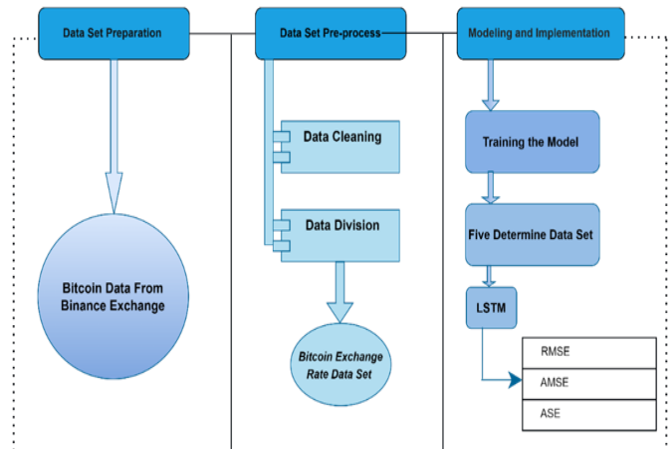


Fig. 1. Proposed LSTM-BTC model.

Long Short-Term Memory (LSTM) and Gated Recurrent Models are examples of RNN extensions (GRU). The problem of remote activities' fading effects in the RNN community is resolved by long short-term memory. It features a transfer that allows you to choose certain actions to be remembered. Additionally, it isn't usually well-established over time and doesn't require as much training. It uses four levels to determine the output before sending the hidden kingdom and the finished product to the cycle after that. To determine if the enjoyment must no longer be counted, "forgetting gates" are present similarly to four layers. Different data may be provided to four levels and forgetting gates for awareness of either short- or long-term memory. When compared to the LSTM model, the GRU or Gated Recurrent Model is seen to be one of the less challenging models since it combines the "forget" and "input" steps into a single step, which only requires the simplest hidden unit.

E. Suggested Model Reasons

- An RNN variant created expressly to address the disappearing gradient issue is the LSTM network. When the erroneous signal backpropagates across several layers of the network and shrinks too much to be meaningful, this issue might arise in RNNs. By employing gates, which manage the information flow over the network, LSTMs get around this issue.
- The redesigned LSTM network suggested in the research enhances the performance of LSTMs by including a control gate and cellular connections to the potential hidden states. This enables the network to discover more intricate correlations between the historical and predicted values of the price of Bitcoin.

- By reducing noise from the data, interpolation fitting and Fourier transform noise reduction techniques help to increase the accuracy of the forecasts.

F. Implementation Details

The implementation details are given in Table II.

TABLE II. IMPLEMENTATION DETAILS

Epoch	Initial Learning Rate	Hardware Resources
100	0.005	Single CPU

V. MATERIALS AND METHODS

A. Simulation and Results

In this Section, analyze the data set first and then optimize the price at various time intervals, including monthly, weekly, daily, hourly, and minutely. Therefore, in the first phase, by observing, the Bitcoin Price with the rate that was in effect at the time for each of the data sets that are used for this study. Fig. 2 is showing that the Proposed LSTM-BTC Bitcoin Price Rate Hourly.

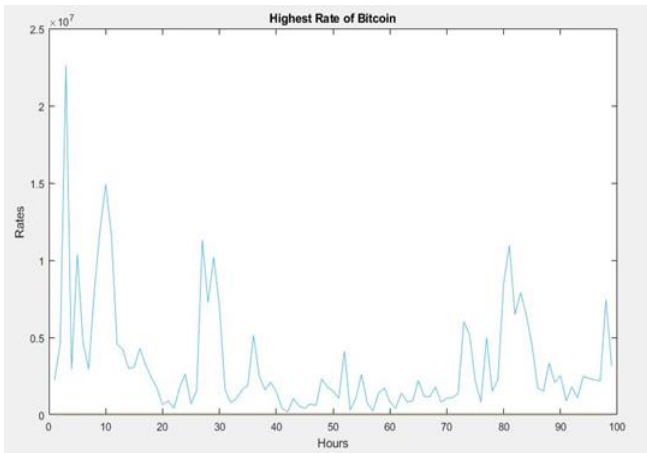


Fig. 2. Proposed LSTM-BTC Bitcoin price rate by hourly.

After that Trained the Model with LSTM on each of the Available Data sets, which will get us the results of the training that are shown below in Fig. 3 to 7.



Fig. 3. Proposed LSTM-BTC trained model with monthly basis.



Fig. 4. Proposed LSTM-BTC trained model with weekly basis data.



Fig. 5. Proposed LSTM-BTC trained model with daily basis data.



Fig. 6. Proposed LSTM-BTC trained model with hourly basis data.



Fig. 7. Proposed LSTM-BTC trained model with minutes basis data.

Determine suitable analytical strategies to study the relationship between the price of bitcoins and other important parameters. After training the model, check the price forecasting results for each outcome in the proposed Model dataset in Fig. 8 to 10.

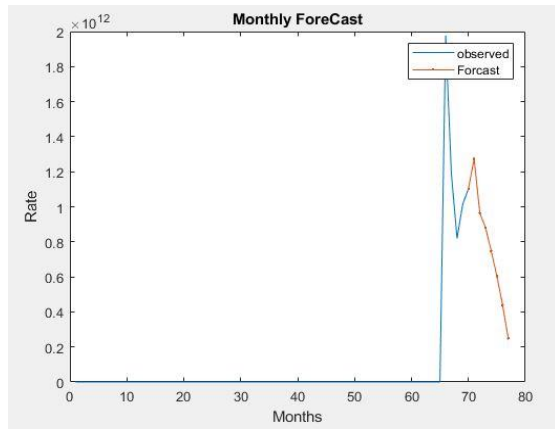


Fig. 8. Proposed LSTM-BTC after training the monthly forecast.

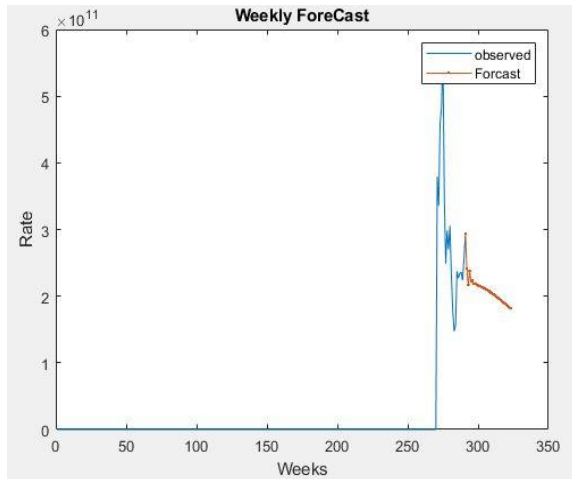


Fig. 9. Proposed LSTM-BTC after training the weekly forecast.

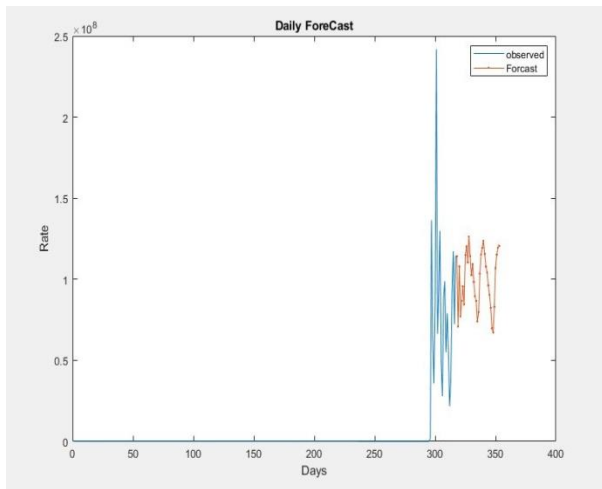


Fig. 10. Proposed LSTM-BTC after training the daily forecast.

After the trained model was applied to each data set, that observed the price forecast with respect to time was zone-specific. After that, examine the updated forecast results using RMSE (Root Mean Square) Value and Graph by applying Validation to the remaining 10% of the data set. It will apply to each one of these data sets as well. Fig. 11 shows the Proposed LSTM-BTC Model Update with Daily Forecast.

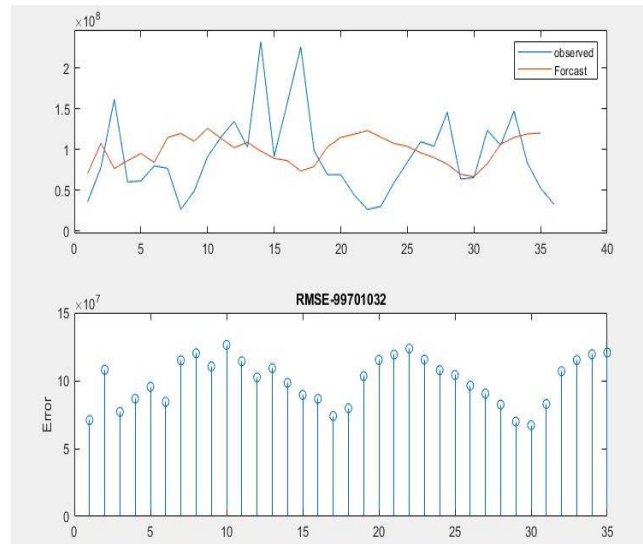


Fig. 11. Proposed LSTM-BTC daily forecast with update.

Before the results' ultimate output, attempt to quantify them using the RMSE, i.e. (root mean square error). Always, RMSE will be higher than or equivalent to MAE. The RMSE scale evaluates a model's capacity to forecast continuous values. To determine whether the margin of error makes sense, the RMSE units are identical to the data units of the dependent variable/target (i.e., if it's in dollars, it's in dollars). The effectiveness of the model improves with decreasing RMSE. Measuring the effectiveness of time series models' short- and long-term predictions is a frequent method of comparison. Utilize the performance measures MAPE (percentage mean absolute error) and RMSE (Root mean square error) to assess the performance of these two models. Utilizing LSTM, these incorrect values were discovered. Fig. 12 to 14 display the RMSE results with respect to months, hours and minutes.

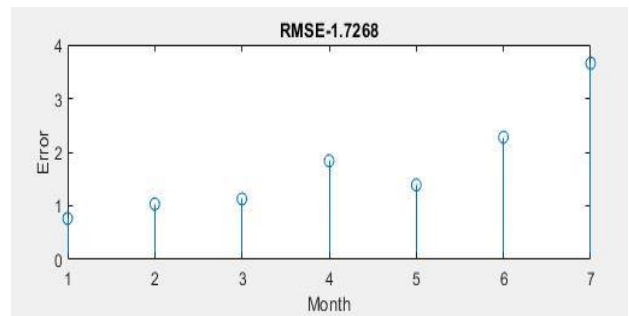


Fig. 12. Proposed LSTM-BTC RMSE of months.

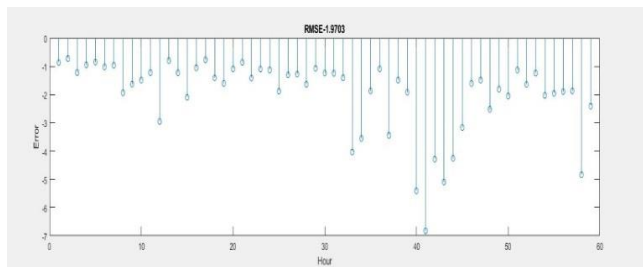


Fig. 13. Proposed LSTM-BTC RMSE of hours.

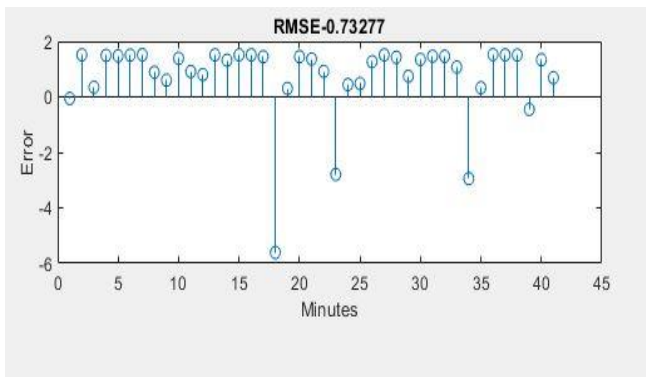


Fig. 14. Proposed LSTM-BTC RMSE of minutes.

After Standardizing the Data Set, the values Mean and Segmentize values are given in Table III.

TABLE III. STANDARDIZED VALUES

Data Set	Mean	Seg.
Monthly	102186464180.311	355077255937.377
Weekly	21721657287.3499	81315350578.9220
Daily	6025642.99833914	24936137.0634950
Hourly	379850.632288345	1768351.56637383
Minutes	33943.5608210427	25640.9457250926

Table IV shows the performance analysis of the proposed LSTM-BTC with respect to Mean, RSME, Absolute Mean Square Error (AMSE), and Absolute Square Error (MAE) for all five datasets: monthly, weekly, daily, hourly and minutes.

TABLE IV. PERFORMANCE ANALYSIS OF PROPOSED LSTM-BTC W.R.T DIFFERENT ERROR AND RESULTS WITH LSTM

Data Set	RSME	AMSE	MAE
Monthly	1.72682328867215	8.4977586e+23	6.6361072e+11
Weekly	0.543983891659373	4.2930577e+22	1.8160258e+11
Daily	0.00989690855714705	1.0233483e+16	67147840
Hourly	0.795046379048935	5.5483145e+12	476020.22
Minutes	0.732772033979149	1.6737144e+09	21938.914

It is also observed that the proposed forecasting model is reliable for a monthly-based dataset with respect to RMSE outputs based on the estimated rate analysis as well as the aforementioned forecasting outcomes.

B. Performance Analysis and Comparison

Overall, the empirical results presented in the previous section support the general hypothesis that “Bitcoin’s coverability rate is determined by various economic and technological determinants over time.” Therefore, rather than using historical exchange rates, higher forecast performance may be attained by exploiting knowledge buried in economic and technical drivers. The results here change by comparing the results here with the final execution of the research paper, so use the latest dataset for this purpose and many conditions change. Table V shows the performance analysis of the proposed method with previously published state-of-the-art approaches.

TABLE V. PERFORMANCE ANALYSIS OF PROPOSED METHOD WITH STATE-OF-THE-ART PREVIOUS PUBLISHED APPROACHES

Approach	Year	Dataset Period	Method	Data Set instances	RMSE
Huang et al. [2]	2021	19/03/2021 to 27/03/2021	LSTM	Day Wise	0.92
Fan Fang et al. [4]	2021	02/07/2018 to 03/07/2018	LSTM	By Hourly	0.82
Awoke et al. [10]	2020	01/01/2014-20/02/2018	LSTM	Day wise	0.092
M.J. Hamayel et al. [11]	2021	01/01/2018 to 30/06/2021	LSTM	Day wise	410.399
Proposed LSTM-BTC	2022	Jan 2021 to March 2022	LSTM	Monthly	1.7268
				Weekly	0.5439
				Daily	0.0098
				Hourly	0.7950
				Minutes	0.7327

VI. CONCLUSION

This study looked into the accuracy of forecasting Bitcoin exchange rates based on technological and economic variables. The DNN-based machine learning model and trained it using historical Bitcoin price data. The study does, however, have certain shortcomings. First, it is unclear how well the model would perform given future data because it was trained on historical data. Second, it is unclear how well the model would generalize to other markets because it was not tested on other cryptocurrencies. Despite these drawbacks, the study significantly adds to the body of knowledge on predicting Bitcoin prices. It thinks that its findings will be valuable to academics, investors, and decision-makers. For upcoming research, look at the use of reinforcement learning and natural language processing as additional machine learning techniques for predicting Bitcoin prices. Test the model's performance on several cryptocurrencies, including Litecoin, Ethereum, and Ripple. Create strategies for increasing the amount of real-time data that Bitcoin price prediction algorithms can use. Research the effects of variables including legislation, public opinion, and technological advancement on the volatility of the Bitcoin price. It is hoped that the effort would stimulate more investigation into this vital subject.

REFERENCES

- [1] Joshi, M., Das, D., Gimpel, K., Smith, a. N.: Movie reviews and revenues: An experiment in text regression. human language technologies. In: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. pp. 293–296 (2010)
- [2] Huang, Xin, and Wenbin Zhang. LSTM-Based Sentiment Analysis for Cryptocurrency Prediction. Vol. 2103, no. 14804v4, 17 Oct. 2021.
- [3] Chen, Wei, et al. “Machine Learning Model for Bitcoin Exchange Rate Prediction Using Economic and Technology Determinants.” International Journal of Forecasting, vol. 37, no. 1, Jan. 2021, pp. 28–43, 10.1016/j.ijforecast.2020.02.008. Accessed 11 Nov. 2021.
- [4] Fan Fang, Waichung Chung, Carmine Ventre, Michail Basios, Leslie Kanthan, Lingbo Li & Fan Wu (2021): Ascertaining price formation in cryptocurrency markets with machine learning, The European Journal of Finance, DOI: 10.1080/1351847X.2021.1908390

- [5] R. Chen and M. Lazer, "Sentiment Analysis of Twitter Feeds for the Prediction of Stock Market"
- [6] Gajardo, G., Kristjanpoller, W.D. & Minutolo, M., 2018. Does bitcoin exhibit the same asymmetric multifractal cross-correlations with crude oil, gold, and DJIA as the euro, Great British pound, and Yen? *Chaos, Solitons & Fractals*, 109, pp.195–205.
- [7] Borges, T.A., Neves, R.N. "Ensemble of Machine Learning Algorithms for Cryptocurrency Investment with Different Data Resampling Methods", *Applied Soft Computing Journal*, Vol. 90, (2020), 106-187. doi 10.1016/j.asoc.2020.106187
- [8] Derbentseva, V. and Babenko, V., 2021. Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices. *International Journal of Engineering*, 34(1).
- [9] Sreekanth Reddy, Lekkala. "A Research on Bitcoin Price Prediction Using Machine Learning Algorithms." *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, vol. 09, no. 22778616, 4 Apr. 2020.
- [10] Awoke, Temesgen. "Bitcoin Price Prediction and Analysis Using Deep Learning Models." *Research Gate*, no. 10.1007/978-981-15-5397-4_63, Oct. 2020.
- [11] Hamayel, Mohammad J., and Amani Yousef Owda. "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM, and Bi-LSTM Machine Learning Algorithms." *AI*, vol. 2, no. 4, 13 Oct. 2021, pp. 477–496, 10.3390/ai2040030.
- [12] Rajua, S M. "Real-Time Prediction of BITCOIN Price Using Machine Learning Techniques and Public Sentiment Analysis." *International Conference*, vol. 00, no. 000, 2021.
- [13] Jagannath, Nishant, et al. "An On-Chain Analysis-Based Approach to Predict Ethereum Prices." *IEEE Access*, vol. 9, no. 0000, 2021, pp. 167972–167989, ieeexplore.ieee.org/document/9650873, 10.1109/ACCESS.2021.3135620. Accessed 8 June 2022.
- [14] Zheng, P. Zhao, K. Huang, and G. Chen, "Understanding the property of long-term memory for the LSTM with attention mechanism," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 2021, pp. 2708–2717.
- [15] Livieris, I., Kiriakidou, N., Stavroyiannis, S. and Pintelas, P., 2021. An Advanced CNN-LSTM Model for Cryptocurrency Forecasting. *Electronics*, 10(3), p.287.
- [16] Livieris, I.E., Stavroyiannis, S.; Pintelas, E.; Pintelas, P. A novel validation framework to enhance deep learning models in time-series forecasting. *Neural Comput. Appl.* 2020, 32, 17149–17167
- [17] Pintelas, E.; Livieris, I.E., Stavroyiannis, S.; Kotsilieris, T.; Pintelas, P. Investigating the Problem of Cryptocurrency Price Prediction: A Deep Learning Approach. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 99–110
- [18] M. Dixon, D. Klabjan and J. H. Bang, "Classification-based financial markets prediction using deep neural networks," *ArXiv*, 2017.
- [19] S. McNally, J. Roche, and S. Caton, "Predicting the price of Bitcoin using machine learning," in *26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*, 2018.
- [20] Ferdiansyah, Ferdiansyah & Othman, Siti & Raja Mohd Radzi, Raja Zahilah & Stiawan, Deris & Sazaki, Yoppy & Ependi, Usman. (2019). An LSTM-Method for Bitcoin Price Prediction: A Case Study Yahoo Finance Stock Market. 206-210. 10.1109/ICECOS47637.2019.8984499.
- [21] Xincheng Zhang, Linghao Zhang, Qincheng Zhou, Xu Jin, "A Novel Bitcoin and Gold Prices Prediction Method Using an LSTM-P Neural Network Model", *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1643413, 12 pages, 2022. <https://doi.org/10.1155/2022/1643413>
- [22] Kjærland, F. et al., 2018. An analysis of Bitcoin's price dynamics. *Journal of Risk and Financial Management*, 11(4), p.63.
- [23] Jiang, X., 2020. Bitcoin price prediction based on Deep Learning Methods. *Journal of Mathematical Finance*, 10(01), pp.132–139.
- [24] Ho, A., Vatambeti, R. & Ravichandran, S.K., 2021. Bitcoin price prediction using machine learning and artificial neural network model. *Indian Journal of Science and Technology*, 14(27), pp.2300–2308.
- [25] Y. Soylemez, "Prediction of gold prices using multilayer artificial neural networks method," *Sosyoekonomi*, vol. 228, no. 46, pp. 271–291, 2020.
- [26] R. Zhang, C. Zhang, and M. Yu, "A similar day-based short-term load forecasting method using wavelet transform and LSTM," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 17, no. 4, pp. 506–513, 2022.
- [27] D. Wu, X. Wang, and S. Wu, "A hybrid method based on extreme learning machine and wavelet transform denoising for stock prediction," *Entropy*, vol. 23, no. 4, pp. 1–30, 2021.
- [28] Borges, T.A., Neves, R.N. "Ensemble of Machine Learning Algorithms for Cryptocurrency Investment with Different Data Resampling Methods", *Applied Soft Computing Journal*, Vol. 90, (2020), 106-187. doi: 10.1016/j.asoc.2020.106187
- [29] Barbon, A. 2019. "Focusing at High Frequency: An Attention-Based Neural Network for Limit Order Books." *Biais, B., P. Hillion, and C. S. Spatt. 1995. "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse." Journal of Finance* 50 (5): 1655–1689.
- [30] Sin E, Wang L. Bitcoin price prediction using ensembles of neural networks. In: *2017 13th International conference on natural computation, fuzzy systems, and knowledge discovery. IEEE. 2017;p. 666–671. doi:10.1109/FSKD.2017.8393351.*

Research on Improving Piano Performance Evaluation Method in Piano Assisted Online Education

Huayi Qi¹, Chunhua She^{2*}

School of Music and Dance, Guizhou Education University, Guiyang 550018, China¹
School of Humanities, Tongren University, Tongren 554300, China²

Abstract—With the continuous progress of science and technology and the popularization of the Internet, online piano education has gradually emerged. This educational model provides piano learning resources and communication platforms through the network platform, so that students can learn piano at home anytime and anywhere. However, there are still some problems in the evaluation method of piano assisted online education, which hinders the development of piano assisted online education. Aiming at the problem that piano assisted online education is difficult to evaluate correctly, this paper proposes to integrate the bidirectional long and short memory network into the instrument digital interface piano performance evaluation model, and to integrate the attention mechanism into the bidirectional long and short memory network, hoping to improve the evaluation accuracy of the model. In the comparison experiment of the evaluation model based on the bidirectional long term memory network, it is found that the accuracy of the bidirectional long term memory network model is 0.91, which is significantly higher than the comparison model. In addition, through the empirical analysis of the model, it is found that the piano online education course integrated with the model can improve students' performance level scores and promote their participation enthusiasm. The above results indicate that the digital interface piano performance evaluation model can not only be used to evaluate digital interface piano performance more accurately, but also to promote the development of online piano education.

Keywords—Short-term memory network; attention mechanism; musical instrument digital interface; online education; piano performance evaluation model

I. INTRODUCTION

Piano performance is a subject full of artistry and technique that places high demands on the performer's musical understanding, skill and expression. In the process of piano learning, effective performance evaluation is crucial for the progress of students [1]. However, traditional methods of evaluating piano performance are often limited by subjective factors, which tend to produce inaccurate and unfair evaluation results [2]. With the continuous progress of science and technology and the popularity of the Internet, piano-assisted online education has gradually emerged [3]. This education model provides piano learning resources and communication platforms through the network platform, so that students can learn at home anytime and anywhere [4]. However, in piano assisted online education, there are still

some problems in the evaluation method, such as unclear evaluation standards and inaccurate evaluation results, which limit students' progress in the learning process [5]. Therefore, the aim of this study is to improve the midi piano performance evaluation method so that it can be better applied in piano assisted online education. Specifically, we will use BiLSTM to analyse the musical characteristics and technical requirements of piano performance, and establish a scientific and objective evaluation system. By analyzing the midi data of students' performances, the accuracy, sound quality, expressiveness and other aspects of performance can be accurately evaluated. This research innovatively introduces the two-way LSTM model into the piano performance evaluation model, so as to realize the automatic evaluation and feedback of students' performance, help students find and correct problems in time, and improve the learning effect. At the same time, the improved evaluation method can also serve as a reference for piano teachers, assist them in teaching and guiding students, and promote the sustainable and healthy development of the piano education industry. This research is mainly divided into five parts. The first part analyzes the status quo of the evaluation model and the LSTM algorithm. The second part describes the construction process of the improved piano performance evaluation model based on LSTM. The third part is the comparative analysis of the performance of the improved algorithm and the evaluation model based on the improved algorithm. The fourth part is the discussion of the results of this study; the fifth part is the summary of the full text.

II. LITERATURE REVIEW

With the boom of computer technology, there are several new methods applied to the evaluation model. Ju's team proposed an evaluation model with the cuckoo algorithm to evaluate the permeability of natural fractures more accurately and found the evaluation model could improve the accuracy of the evaluation of the permeability of natural fractures [6]. Chao et al. proposed a wheat yield evaluation model based on a simple algorithm to improve the accuracy of the winter wheat yield estimation model and used the model for empirical analysis, which was found to be of great relevance in estimating not only the yield of wheat accurately, but also the biomass and yield future sensing data [7]. Lai's team has improved the reliability evaluation model for public transportation routes and proposed an information entropy based reliability evaluation model for public transportation

routes, which has been empirically analyzed. The experimental results found that the overall reliability of this model is higher than that of traditional evaluation models [8]. Li and Sun proposed a cloud computing-based English-speaking quality evaluation model to improve the precision of the English teaching quality evaluation model, and the model for comparative experimental analysis, and the results showed this method can evaluate the teaching quality of spoken English with accurate evaluation results [9]. Ding et al. proposed a model for evaluating the timeliness of online instruction with intelligent learning to accurately assess the quality of online instruction, and this study used statistical feature analysis methods to statistically analyze and robustness test the model, and the results showed this model has high credibility in evaluating the timeliness of online teaching, which can increase the precision of online teaching quality assessment [10].

With the wide application of neural network technology, LSTM neural network has been applied in many fields. Guo's team proposed an LSTM-based path planning algorithm for mobile robots and conducted simulation experiments to show that the algorithm not only improves the computational speed compared with similar algorithms but also can adapt to an environment with many obstacles [11]. Li et al. designed a hybrid model with LSTM neural network for the prediction of monthly runoff to resolve low precision of water utilization prediction models. The model was empirically analyzed and found the prediction model was more accurate than similar models and provided a reliable basis for the full utilization of water resources [12]. Geng proposed a deep learning architecture based on LSTM neural networks to grasp the nature of patents more quickly and intuitively, used the architecture to accurately simplify the patent text, and found the architecture can increase the efficiency of researchers in mastering patents [13]. Yao's team proposed a reinforcement learning network with an LSTM network to predict the tool life to increase the machining quality of tools and the productivity of the tool automation system, and the empirical analysis of the proposed network showed that the accuracy of the network in predicting the tool life was is greater than that of traditional prediction methods, which is important for improving the machining quality of tools [14]. Altuve and Hernandez proposed a heart rhythm identification model based on the bidirectional LSTM technique for the problem of insufficient measures for early cardiovascular disease detection, and the proposed model was subjected to empirical analysis, and the results showed that the model can accurately identify heart rhythm identification and to discriminate, providing technical support for early detection of cardiovascular diseases and prompt action to protect people's health [15].

Through the above research, it can be found that the current application range of LSTM neural networks is wide, and there are many methods applied to evaluation models. Through the specific analysis of the above studies, it is found that there is little research that combines LSTM neural network with piano performance evaluation model at present. In this study, the bidirectional LSTM neural network is applied to the MIDI piano performance evaluation model,

which is quite different from previous research. It is expected to improve the accuracy of MIDI piano performance evaluation model with LSTM neural network, and to provide technical basis for piano assisted network education.

III. CONSTRUCTION OF IMPROVED PIANO PERFORMANCE EVALUATION MODEL BASED ON LONG-TERM MEMORY NETWORK

A. Long-term Memory Network Based on Recurrent Neural Network

Traditional RNNs are prone to the problem of gradient disappearance, and the analysis is not accurate when encountering long sequence-type data [16]. The LSTM, as a special recurrent NN, can resolve the question well by improving the hidden layer structure [17]. Some special "gate" structures are added to the neurons of each layer of the LSTM neural network. The purpose of this is to make the error not all attributed to the current neuron in the propagation process, but some of it directly through the "gate" structure, only in this way can the error be well directed to the next layer to avoid the phenomenon of disappearing gradient. This also leads to better convergence [18].

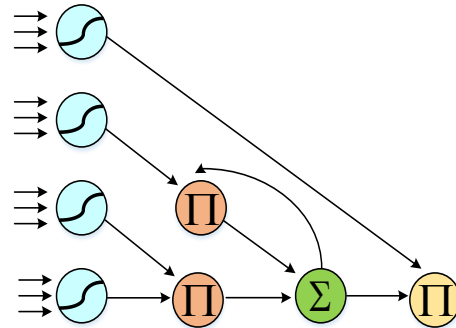


Fig. 1. Memory structure of the LSTM model.

Fig. 1 shows the structure of the LSTM neural network. Compared with the RNN NN, LSTM NN improves the performance by changing the structure so that it can handle long-time series data. This is demonstrated by the introduction of three gating structures, namely, forgetting gates, input gates and output gates. The expression of the forgetting gate f_t is shown in equation (1).

$$f_t = \sigma(W_f h_{t-1} + U_f x_t + b_f) \quad (1)$$

In equation (1), f_t represents the calculation rule of the forgetting gate at the moment of t , W_f and U_f are the parameter matrices, b_f represents the bias term and σ represents the Sigmoid activation function. equation (2) is an expression for the Sigmoid function.

$$\delta(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

From equation (2), it can be seen that the Sigmoid function makes the output value of the forgetting gate lie between [0,1], which can make the data better aggregated in the transmission process of the network. So, the Sigmoid activation function is

also chosen for the calculation in the output and input gates. The function of the input gate of the gating structure is a one-time filtering of the current input information to determine what percentage of the current information is added to the current cell state, and the input gate i_t is calculated as shown in equation (3).

$$i_t = \sigma(W_i h_{t-1} + U_i x_t + b_i) \quad (3)$$

In equation (3), i_t represents the calculation rule of the input gate at the moment of t , σ represents the Sigmoid function, W_i and U_i are the parameter matrices, and b_i represents the bias term. When the current new information is received, some information from the previous moment is superimposed with a certain probability to form the new input information. The expression of the new memory \tilde{C}_t is shown in equation (4).

$$\tilde{C}_t = \tanh(W_c h_{t-1} + U_c x_t + b_c) \quad (4)$$

In equation (4), x_t represents the input information at the time of t , h_{t-1} represents the hidden state at the time of $t-1$, W_c and U_c are the parameter matrices, and b_c represents the bias term. \tanh function is an activation function, also called hyperbolic tangent function, which is mainly used in neuron state and output calculation, and its calculation equation is shown in equation (5).

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

From equation (5), the output of the \tanh activation function lies between $[-1,1]$ and, in particular, when the input value is 0, the output of the function is also 0. Using the \tanh activation function is equivalent to adjusting the mean of the input values to 0, which facilitates subsequent processing. When a new message is received, the cell state is updated by multiplying x_t and \tilde{C}_t to the new cell state, the forgetting gate and the input gate change the current cell state C_t by probabilistically selecting the previous moment and the current message, the cell state is updated from the original C_{t-1} to the current C_t process, and the final memory C_t is specified in equation (6).

$$C_t = C_{t-1} \square f_t + i_t \square \tilde{C}_t \quad (6)$$

In equation (6), denotes the cell state at the moment of C_{t-1} $t-1$, f_t denotes the state of the forgetting gate at the moment of t . i_t represents the input gate state at time t . The final memory of the current cell state is obtained by adding the

information passed by the old cell plus the filtered content of the new information C_t . The output gate extracts information from the current cell state, and the output gate o_t expression is given in equation (7).

$$o_t = \sigma(W_o h_{t-1} + U_o x_t + b_o) \quad (7)$$

In equation (7), o_t represents the computation rule of the output gate at the time of t , x_t represents the input information at the time of t , h_{t-1} represents the hidden state at the time of $t-1$, W_o and U_o are the parameter matrices, and b_o represents the bias term. The extracted information is used to generate the hidden state h_t and the expression of the hidden state at the time of t is shown in equation (8).

$$h_t = o_t \square \tanh(C_t) \quad (8)$$

In equation (8), the final memory state of the cell at time t , denoted as C_t is input into the hyperbolic tangent function for calculation. Subsequently, the output gate, denoted as o_t , is combined with it through an operation represented as \square resulting in the extraction of the information component h_t . The equation (9) can be obtained by associating equations (4), (6) and (8).

$$h_t = o_t \square \tanh(C_{t-1} \square f_t + i_t \square \tanh(W_c h_{t-1} + U_c x_t + b_c)) \quad (9)$$

From equation (9), h_t is the size of the hidden state at the moment of t . W_c is the main cause of the gradient disappearance inside the traditional recurrent NN, while here, when the forgetting gate f_t is opened, the gradient of C_t can be effectively passed to the cell state C_{t-1} at the previous moment; so by adding the gating structure on top of the traditional recurrent NN.

B. Bidirectional Long and Short Time Memory Network and Improvement Method

LSTM can effectively improve the gradient disappearance and gradient explosion problems though. However, it can only process data in one direction in the time series problem, which often ignores future information [19-20]. In contrast, bidirectional LSTM networks can propagate not only forward but also backwards, and this combination of combining past information with future information is similar to the human appreciation of music [21-22]. Using this approach to train the MIDI piano playing evaluation model, more reliable parameters can be obtained and the piano performance evaluation accuracy can be improved. Fig. 2 shows the structure of a bidirectional recurrent NN.

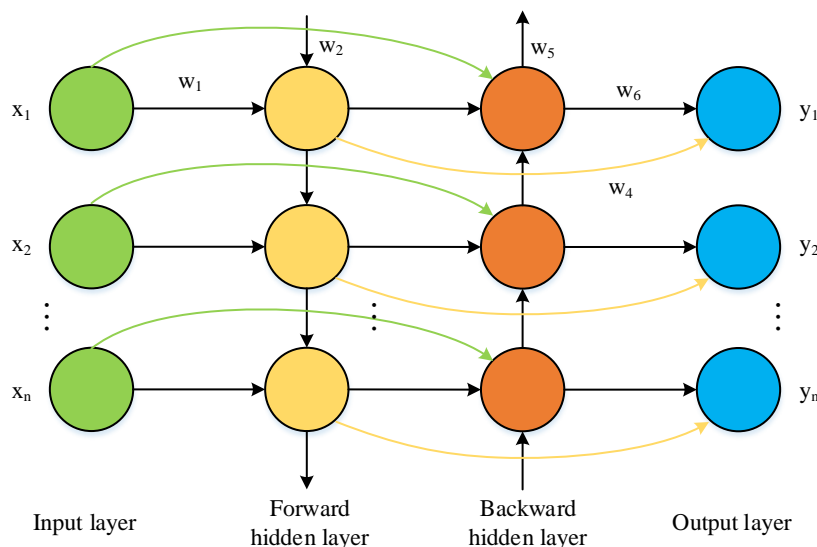


Fig. 2. Structure of bidirectional recurrent neural network.

In Fig. 2, the bidirectional recurrent NN consists of two recurrent neural networks connected backwards and use the tanh function as the activation function, and the output layer receives information not only from forward propagation, but also from backward propagation, and there is information exchange between the forward hidden layer and the backward hidden layer. In Fig. 2, $w_1 - w_6$ are six different ways of information transfer, which have the same weights in the information transfer process. It is difficult to establish the dependency between the current information and the remote information in the bidirectional recurrent neural network because of the gradient descent. Therefore, the neuron module in the hidden layer of the bidirectional recurrent NN is replaced with the LSTM network module to become a bidirectional LSTM model. The bidirectional LSTM model can not only effectively reduce the probability of gradient explosion, but also maintain the dependency relationship between the current value and the remote value. This way of processing both past and future information is consistent with the traditional music evaluation method, and can better judge the expression and coherence of music so that the MIDI piano performance can be accurately evaluated. To better apply the bidirectional LSTM network to the MIDI piano playing evaluation model, the study combines the attention mechanism and the Softmax function, which is mainly to resolve the difficulty of remote information dependence, and its expression is shown in equation (10).

$$\alpha = \frac{\exp(e_{t,i})}{\sum_{j=1}^N \exp(e_{t,j})} \quad (10)$$

In equation (10), $\exp(e_{t,i})$ denotes the output value of the bidirectional LSTM model. After the input music information is passed through the attention mechanism layer, it is then classified using the Softmax function, which is normalized to obtain the probability of which category the MIDI music belongs to. The Softmax function is a nonlinear function that maps the output of multiple neurons to a vector of real

numbers in the interval (0,1), and the sum of all elements in the real vector is 1. Its expression is shown in equation (11) is shown.

$$S(x_j) = \frac{e^x j}{\sum_{k=1}^k e^x k}, j = 1, 2, \dots, K \quad (11)$$

In equation (11), $S(x_j)$ denotes the value of the i -th dimension of the feature vector and k denotes the number of categories. Equation (12) is the equation for the MIDI piano playing classification label.

$$\hat{m} = \arg \max S(x_j) \quad (12)$$

There is a very simple functional relationship between the gradient of the cross-entropy function and the output value of Softmax, which can save a lot of time in the gradient calculation and make the calculation faster and more stable. Therefore, the loss function for the training of the MIDI music evaluation prediction model is chosen as the categorical cross-entropy function, whose functional expression is shown in equation (13).

$$L = -\sum_{j=1}^T y_j \log s_j \quad (13)$$

In equation (13), s_j denotes the estimated probability of each category in the classification, and T is the number of categories in the classification. The study improves the precision of MIDI piano playing evaluation by incorporating the attention mechanism and Softmax function in the bidirectional LSTM network, and better provides technical support for the development of piano online education.

C. Design of Piano Performance Evaluation Model Based on Two-way Long and Short Time Memory Network

To perform accurate evaluation of MIDI piano performance, the study proposes to implement a bidirectional LSTM neural network model with an attention mechanism on

the framework of Spark and Deeplearning4J to evaluate MIDI piano performance with this model. The model mainly consists of a data acquisition module, data pre-processing module and music evaluation classification module, and its specific framework diagram is shown in Fig. 3.

From Fig. 3, the MIDI piano playing evaluation model is mainly divided into three modules: data acquisition, data pre-processing and music evaluation classification. The data pre-processing module mainly filters the original data that are not suitable for model training or transforms them into a matrix suitable for model training, and divides the data into a training set, validation set, and test set; the data acquisition module uses the acquisition tool to migrate data to the storage

system; the model is primarily trained on pre-processed data in the classification module. The parameters of the model are then adjusted in real-time in response to the training results to obtain model parameters with better evaluation effects. The test set data are then trained to obtain the rating prediction results for the purpose of classification by the model after the parameters have been adjusted. In the evaluation of MIDI piano performance is usually evaluated for multiple MIDI music, and the evaluation results are divided into five grades: excellent, good, moderate, poor, and bad. The study uses the bidirectional LSTM NN and attention mechanism layer in deep learning to classify by softmax function to evaluate MIDI piano performance more accurately.

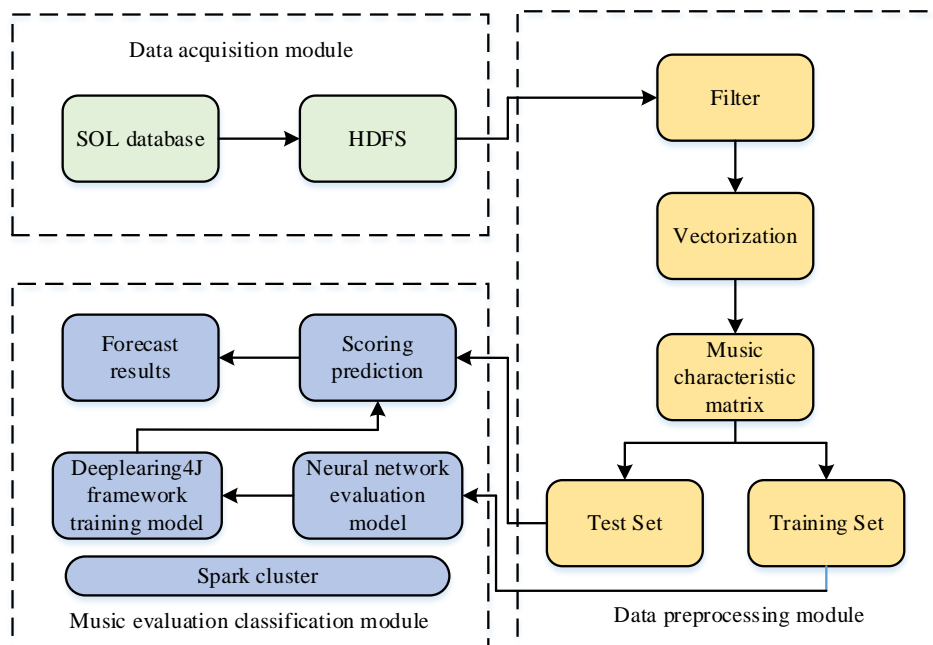


Fig. 3. MIDI-Piano Evaluation Model Framework.

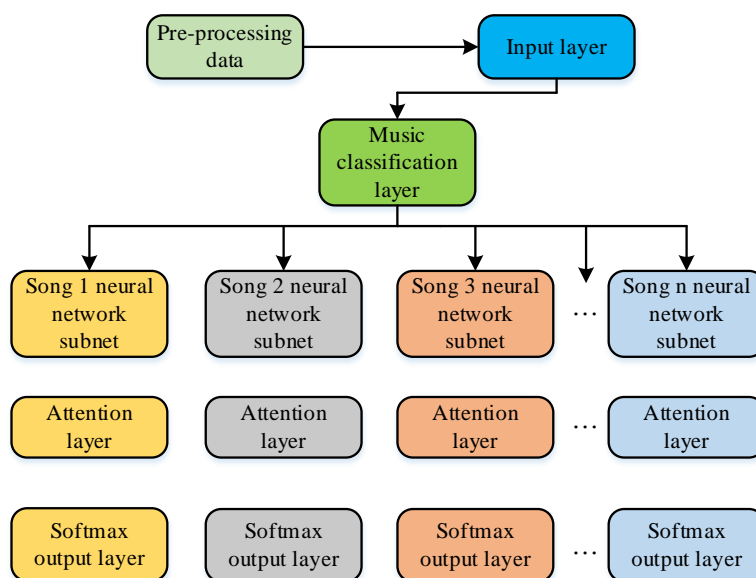


Fig. 4. Bidirectional LSTM neural network evaluation model.

Fig. 4 shows the structure of the bidirectional LSTM NN evaluation model, in which several subnet models are included, and each subnet model needs to be trained separately to ensure that each subnet model is evaluated for a specific piano repertoire, thus realizing the function of evaluating multiple piano repertoires. In the evaluation process of the subnet models, the differences in the training level of each track are used to obtain the feature matrix, and they are classified into one of five categories: excellent, good, moderate, poor, and bad according to the classification algorithm. In each subnet model mainly consists of an input layer, a bidirectional LSTM layer, an attention mechanism layer, and an output layer. The role of the input layer is to receive the difference sample tracks, obtain the input feature matrix through data preprocessing, and then input to the bidirectional LSTM hidden layer, and after the attention mechanism layer, finally, the evaluation is derived by the softmax function in the output layer. In this way, the evaluation results of MIDI piano performance are obtained and used to promote the development of piano online education and achieve the healthy development of piano online education.

IV. COMPARATIVE ANALYSIS OF MODEL PERFORMANCE AND EMPIRICAL EFFECT ANALYSIS

A. Experimental Results Analysis of Model Performance Comparison

To verify the function of the MIDI piano playing evaluation with the bi-directional LSTM model proposed in the study, the BP, RNN, LSTM, and bi-directional LSTM models with a single hidden layer were compared in the experiments, and the model accuracy, precision, recall, and F1 values were compared as the comparison indexes. Accuracy refers to the percentage of the number of samples correctly classified by the classifier to the total number of samples, reflecting the situation where the classifier correctly identifies each sample. The calculation equation is shown in Equation (14).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (14)$$

In equation (14), TP represents true positive; TN represents true negative, FP and FN represent false positive and false negative. Accuracy refers to the number of true cases in the sample where the prediction result is positive,

and the calculation equation is shown in Equation (15).

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

Recall refers to the percentage of positive predictions, with the equation shown in Equation (16).

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

In order to better evaluate the performance of the classifier, the precision and the recall rate are called the measure, and the calculation equation is shown in Equation (17).

$$F_\alpha = \frac{(1 + \alpha^2) * Precision * Recall}{\alpha^2 * Precision + Recall} \quad (17)$$

In equation (17), α is a non-negative real number. When α is 1, it is F1, which is the harmonic mean of precision and recall.

The Google Open Image dataset is used as a test set to compare the accuracy of the four models. The Google Open Image dataset contains 1.9 million images, 600 species and 15.4 million bounding box annotations. It is currently the largest dataset with object location annotation information. The results of the accuracy comparison of the four algorithms are shown in Fig. 5.

Fig. 5 shows the accuracy curves of the four different models compared with epochs, from which the accuracy curves of the bidirectional LSTM model are higher than those of the other three models, and the accuracy curves show an increasing trend with the increase of the number of iterations. The accuracy curve of the two-way LSTM model has a maximum value of 0.91, which is higher than the LSTM model (0.73), the BP model (0.46) and the RNN model (0.38), and the performance of the two-way LSTM model is better than the three comparison models in terms of accuracy dimension. The Google Open Image dataset was used as the training and test set to perform two comparison experiments on the four models, and the results of the four models in five different classification categories were compared. The accuracy rate curves of the four models in the two comparison experiments are shown in Fig. 6.

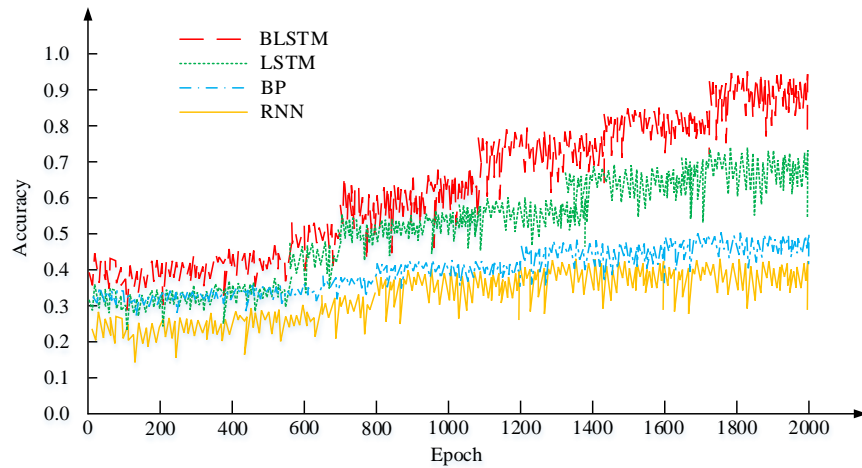


Fig. 5. Comparison of accuracy of different models.

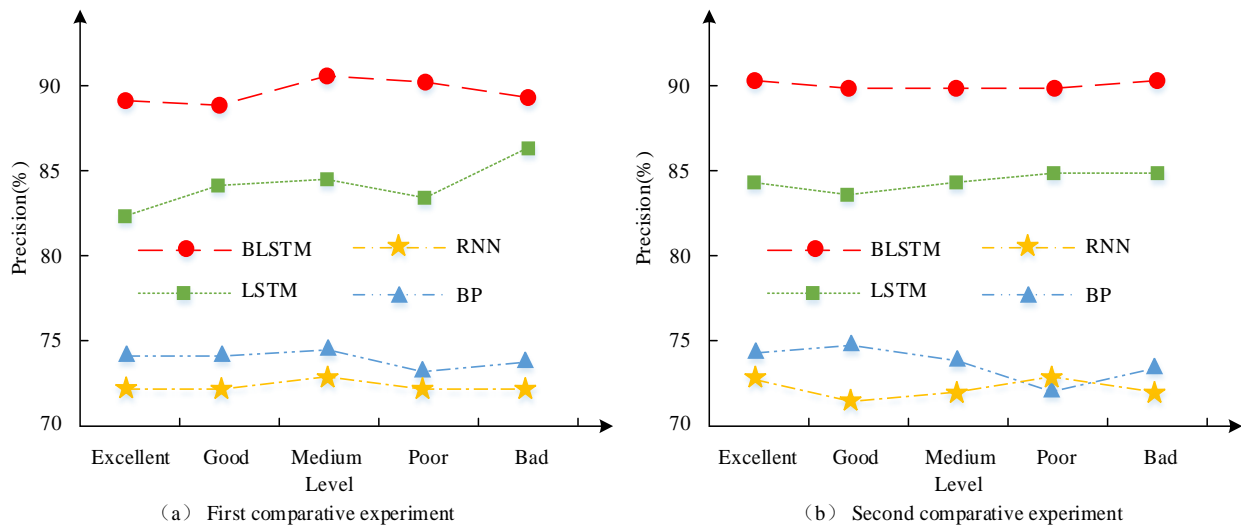


Fig. 6. Accuracy curve of two comparative experiments.

Fig. 6 shows the precision rate curves of the four models in the five classification categories. Fig. 6(a) shows the results of the first comparison experiment. From Fig. 6(a), the precision curves of the bidirectional LSTM model are higher than those of the other three models, and its average precision rate is 88.7 % in the datasets of the five different categories, which is higher than that of the LSTM (83.8%), the BP (73.9%) and RNN (72.6%). Fig. 6(b) shows the results of the second comparison experiment. From Fig. 6(b), the precision curve of the bidirectional LSTM model is higher than the other three models, and its average precision rate in the dataset of five different categories is 89.3%, which is higher than 84.1% of the LSTM model, 73.8% of the BP model, and 72.9% of the RNN model. The above results indicate that the two-way LSTM model has the best performance in terms of the dimension of precision rate. Fig. 7 shows the recall curves of

the four models.

Fig. 7 shows the recall curves of the four models in the five classification categories. Fig. 7 (a) shows the results of the first comparison experiment, from which the recall curve of the bidirectional LSTM is higher than the other three models, and its average recall rate in the five different categories 88.8% is higher than that of LSTM (79.5%), BP (68.4%) and the RNN model (63.4%). Fig. 7(b) shows the results of the second comparison experiment. From Fig. 7(b), the recall curve of the bidirectional LSTM model is higher than the other three models, and its average recall rate in the dataset of five different categories is 89.2%, which is higher than LSTM (79.8%), BP (68.8%), and RNN (63.6%). The above results indicate that the two-way LSTM model has the best performance in terms of the dimension of recall rate. Fig. 8 shows the F1 value curves of the four models.

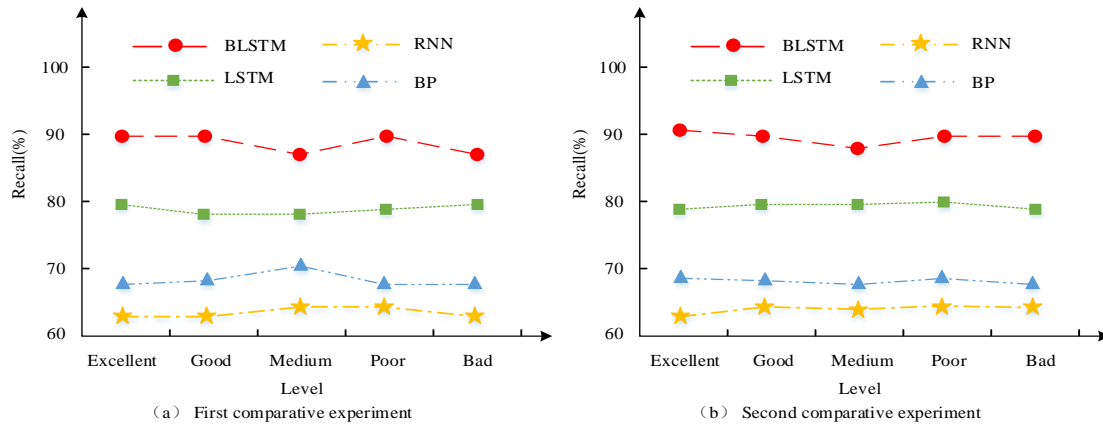


Fig. 7. Recall rate curve of two comparative experiments.

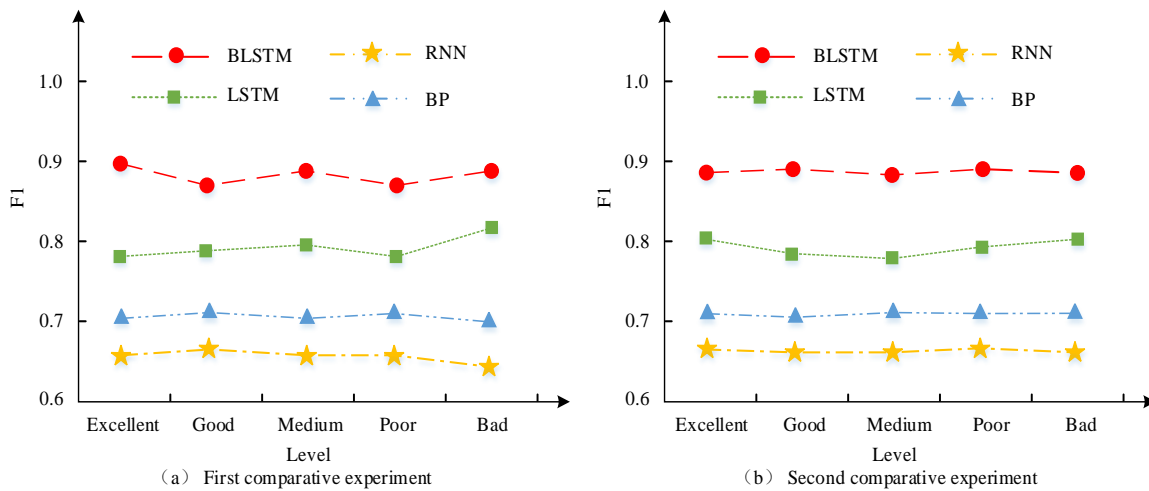


Fig. 8. F1 value curve of two comparative experiments.

Fig. 8 shows the F1 value curves of the four models in the five classification categories. Fig. 8(a) shows the results of the first comparison experiment, from which the F1 value curve of the two-way LSTM is higher than the other three models, and its average F1 value in the five different categories is 0.88, which is higher than LSTM (0.79), BP (0.70) and RNN (0.65). The above results show that the bidirectional LSTM model has a higher F1 value curve than the other three models, and its average F1 value is 0.89 in the data set of five different categories, which is higher than LSTM (0.78), BP (0.71) and RNN (0.66). In conclusion, the overall performance of the bidirectional LSTM model is better than that of the LSTM model, BP model and RNN model, and using this model to evaluate MIDI piano playing can improve the accuracy of the evaluation model.

B. Effect Analysis of Evaluation Model Based on Bidirectional Long Term Memory Network in Piano Online Education

In addition to testing the function of the proposed model, the study also analyzed the actual teaching effects of its integration into piano online education. Students with the same basic information were divided into two groups, and the

experimental group was taught with the piano online education course integrated with the model, while the control group was taught with the regular piano online education course. The results of students' performance level score and willingness to participate in the course for both groups are shown in Fig. 9. Both evaluation indexes have a full score of 10.

As can be seen in Fig. 9(a), the performance score of the experimental group was higher than control group in all five groups, and the average performance level score of the experimental group was 8.9, much higher than control group, which was 7.3. As can be seen in Fig. 9(b), the willingness to participate in lessons was higher than control group in all five groups, and the average willingness to participate in lessons of the experimental group was 8.7, much higher than control group, which was 7.5. The average willingness score of the experimental group is 8.7, which is much higher than the 7.5 score of the control group. These results indicate that the LSTM-based MIDI piano performance evaluation model can effectively improve academic piano performance and willingness to participate in lessons. In addition, Table I shows the results of the questionnaire on students' subjective emotional experiences.

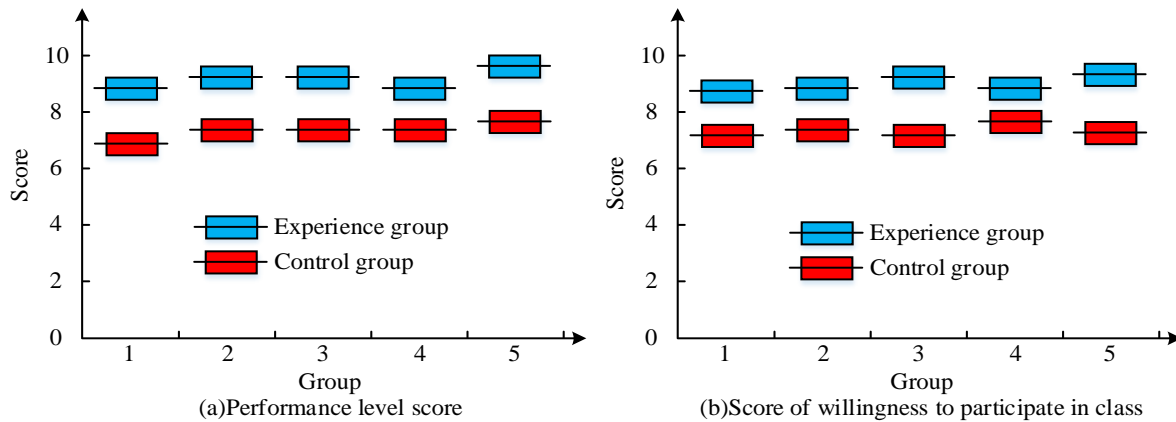


Fig. 9. Performance score and willingness to participate.

TABLE I. SUBJECTIVE MOOD SURVEY RESULTS

/	Experimental class	Control class	T	P
Happiness	8.63±0.92	7.56±0.71	1.12	<0.0001
Psychological annoyance	6.57±0.57	7.33±0.69	-0.56	<0.0001
Feeling of fatigue	7.25±0.74	8.13±0.81	-0.47	<0.0001
Initiative	9.17±0.93	7.58±0.83	0.62	<0.0001
Sense of gain	8.85±0.85	7.32±0.76	0.73	<0.0001
Self-confidence	8.98±0.94	7.62±0.82	0.51	<0.0001
Fulfilment	9.08±1.02	8.03±0.79	0.66	<0.0001

Table I demonstrates that the experimental group students' happiness, motivation, acquisition, self-confidence, and achievement scores were higher than those of the control group students, demonstrating that the piano performance evaluation model proposed in the study can effectively improve students' performance in piano online education courses the positive subjective emotional experience. Additionally, Table I demonstrates that the piano performance evaluation model can lessen students' negative subjective emotional experiences. In conclusion, the piano performance evaluation model applied to the piano online education course can improve the overall teaching quality of the physical education course, promote students' positive emotions in class, and can improve students' piano performance level.

V. DISCUSSION

With the continuous improvement of people's material living standards, their awareness of spiritual needs is also deepening, which promotes the second rapid development of piano education. Whether it is entering art schools for collective learning or seeking individual learning opportunities for piano lessons, the demands on piano education professionals are increasing, and the number of traditional piano teachers in China is seriously insufficient, making the contradiction between supply and demand difficult to solve in the short term. In addition, due to the lack of formal teaching and training institutions and teachers, many piano students acquire piano knowledge and skills through self-study or entrust others with their practice. In the field of piano education, the commission received by traditional teachers is

generally high, and the price of wooden pianos is also very high, which makes it difficult for ordinary families to afford them, and the high cost has become an important obstacle to the development of piano education. In addition, due to the large population in our country, parents and children are the only children, which leads to the outdated concept of family education, the lack of scientific and systematic guidance, so that piano education has been in the "low age" stage for a long time. Therefore, in the process of promoting piano education, reducing the cost of learning has become an essential trend. MIDI is a protocol proposed in the early 1980s to solve the communication between electroacoustic instruments. With the development trend of "network + education", MIDI piano has gradually become an indispensable tool in piano education.

In this study, the MIDI piano performance evaluation model based on bidirectional LSTM was compared with BP, RNN and LSTM models. From the perspective of the accuracy dimension, the accuracy of the bidirectional LSTM model is higher than 0.73 of the LSTM model, 0.46 of the BP model and 0.38 of the RNN model. Compared to the three comparison models, the bidirectional LSTM model has a better performance. Similar results were found in the study by Dandil's team [23]. From the accuracy dimension, in the first comparison experiment, the accuracy of the bidirectional LSTM model is higher than 83.8% of the LSTM model, 73.9% of the BP model and 72.7% of the RNN model. In the second comparison experiment, the accuracy of the bidirectional LSTM model is higher than 84.1% of the LSTM model, 73.8% of the BP model and 72.9% of the RNN model, and the performance of the bidirectional LSTM model is the best

compared to the three comparison models. This is similar to the conclusion obtained by Xie et al. [24]. From the dimension of recall rate, in the first comparison experiment, the recall rate of the bidirectional LSTM model is higher than 79.5% of the LSTM model, 68.4% of the BP model and 63.4% of the RNN model. In the second comparison experiment, the recall rate of the bidirectional LSTM model is higher than 79.8% of the LSTM model, 68.8% of the BP model and 63.6% of the RNN model, and the performance of the bidirectional LSTM model is the best compared to the three comparison models. Tao's team also found similar results [25]. From the perspective of F1 value dimension, in the first comparison experiment, F1 value of the bidirectional LSTM model among the four models was higher than 0.79 of the LSTM model, 0.70 of the BP model and 0.65 of the RNN model. In the second comparison experiment, the F1 value of the bidirectional LSTM model is higher than 0.78 of the LSTM model, 0.71 of the BP model and 0.66 of the RNN model. The performance of the bidirectional LSTM model is the best among the three comparison models. This is in line with the conclusion of Jian et al [26].

In addition to testing the performance of the proposed evaluation model, this study also analyzes the actual teaching effect of integrating it into online piano education. The experimental group took the piano online education course integrated with the model for teaching, and the control group took the conventional piano online education course for teaching, and the score of piano playing level and students' subjective experience were used as evaluation indicators for comparative analysis. Among the five groups, the score of the performance level of the experimental group was 8.9 points higher than that of the control group, and the score of the experimental group's willingness to participate in the class was 8.7 points higher than that of the control group. It can be seen that the MIDI piano performance evaluation model based on LSTM can effectively improve the academic piano performance level and the willingness to participate in the class. This result is consistent with the research findings of Bao's team [27]. In addition, by conducting a questionnaire on students' subjective emotional experience, the scores of happiness, initiative, sense of gain, self-confidence and achievement of students in the experimental group were all higher than those of students in the control group, indicating that the piano performance evaluation model proposed in this study can effectively improve students' positive subjective emotional experience in online piano education courses.

VI. CONCLUSION

In this study, performance comparison experiments were conducted on the proposed two-way LSTM performance evaluation model, and the results showed that the accuracy rate, accuracy rate, recall rate and F1 value of the model were 0.91, 88.7%, 88.8% and 0.88, respectively, which were higher than the other three comparison models. In addition, in the empirical analysis of the model, it is found that the model can not only improve the performance level and enthusiasm of the students, but also promote the positive subjective emotions of the students. In conclusion, the improved MIDI piano performance evaluation model based on two-way LSTM is not only accurate, but can also promote the development of online

piano education. This research innovatively integrates the two-way LSTM network model into the piano performance evaluation model, which not only improves the evaluation accuracy of the piano performance evaluation model, but also improves the integration degree of deep learning and piano education, making up for the lack of integration between the current piano education and deep learning. However, there are still shortcomings in the process of this study, such as crude evaluation and classification methods and unreasonable selection of data sets. In the future, it is necessary to find effective evaluation and classification methods, and to obtain more closely related data sets through self-preprocessing for comparison experiments in order to obtain more accurate results.

ACKNOWLEDGMENT

The research is supported by this paper is the phased result of the 2022 Annual Scientific Research Fund Project of "Guizhou Education University" (No. 2022BS001).

REFERENCES

- [1] Boccardo G, Millo F, Piano A, Arnone L, Manelli S, Fagg S, Gatti P, Herrmann O, Queck D, Weber J. Experimental investigation on a 3000bar fuel injection system for a SCR-free non-road diesel engine. *Fuel*, 2019, 243:342-351.
- [2] Bastepe-Gray S, Riley M A, Klotchkov N, Supnekar J, Filippi L, Raghavan P. Ecology of Musical Performance as a model for evaluation and treatment of a musician with a playing related musculoskeletal disorder: A case report. *Journal of Hand Therapy*, 2021, 34(2):330-337.
- [3] Nakamura E, Saito Y, Yoshii K. Statistical learning and estimation of piano fingering. *Information Sciences*, 2020, 517:68-85.
- [4] Filipovi N, Brdar S, Mimi G, Marko O, Crnojevi V. Regional soil moisture prediction system based on Long Short-Term Memory network. *biostystems Engineering*, 2022, 213:30-38.
- [5] Wu J, Tang J, Zhang M, Di J, Hu L, Wu X, Liu G, Zhao J. PredictionNet: a long short-term memory-based attention network for atmospheric turbulence prediction in adaptive optics. *applied optics*, 2022, 61(13):3687-3694.
- [6] Ju Y, Dong J, Gao F, Wang J. Evaluation of Water Permeability of Rough Fractures Based on a Self-affine Fractal Model and Optimized Segmentation Algorithm. *Advances in Water Resources*, 2019, 129:99-111.
- [7] Chao, Zhang, Jianguai, Liu, Taifeng, Dong, Jiali, Shang, Min, Tang. Evaluation of the Simple Algorithm for Yield Estimate Model in Winter Wheat Simulation under Different Irrigation Scenarios. *Agronomy Journal*, 2019, 111(6):2970-2980.
- [8] Lai Y, Ma Z, Xu S, Easa S M. Information entropy evaluation model of bus-line reliability considering the combination of bus stops and bus travel time. *Canadian Journal of Civil Engineering*, 2022, 49(1):64-72.
- [9] Li H, Sun S. Research on evaluation model of oral English teaching quality based on cloud computing. *international journal of continuing engineering education and life-long learning*, 2020, 30(4):363-380.
- [10] Ding X, Salam Z A, Lv W. Research on Timeliness Evaluation Model of Online Teaching Based on Intelligent Learning. *international Journal of Continuing Engineering Education and Life-Long Learning*, 2021, 31(2):263-275.
- [11] Guo N, Li C, Wang D, Song Y, Gao T. Local Path Planning of Mobile Robot Based on Long Short-Term Memory Neural Network. *Automatic Control and Computer Sciences*, 2021, 55(1):53-65.
- [12] Li B, Sun G, Liu Y, Wang W, Huang X. Monthly Runoff Forecasting Using Variational Mode Decomposition Coupled with Gray Wolf Optimizer-Based Long Short-term Memory Neural Networks. *Water Resources Management: An International Journal, Published for the European Water Resources Association (EWRA)*, 2022, 36(2):2095-2115.

- [13] Geng B. Text segmentation for patent claim simplification via Bidirectional Long-Short Term Memory and Conditional Random Field. *Computational Intelligence*, 2022, 38(1):205-215.
- [14] Yao J, Lu B, Zhang J. Tool Remaining Useful Life Prediction Using Deep Transfer Reinforcement Learning Based on Long Short-Term Memory Networks. *The International Journal of Advanced Manufacturing Technology*, 2021, 118(3):177-1086.
- [15] Altuve M, Hernandez F. Multiclass Classification of Cardiac Rhythms on Short Single Lead ECG Recordings using Bidirectional Long Short-Term Memory Networks. *IEEE Latin America Transactions*, 2021, 19(7):1207-1216.
- [16] Yu X, Li D, Shen Y. Forecasting Stock Index Using a Volume-Aware Positional Attention-Based Recurrent Neural Network. *International Journal of Software Engineering and Knowledge Engineering*, 2021, 31:1783-1801.
- [17] Liu X, Lang L, Zhang S, Xiao J, Fan L, Ma J, Chong Y. Intelligent fault diagnosis of medical equipment based on long short term memory network. *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi*, 2021, 38(2):361-368.
- [18] Salman U, Rehman S, Alawode B, Alhems M. Short term prediction of wind speed based on long-short term memory networks. *fine transactions*, 2021, 49(3): 643-652.
- [19] Hsieh C, Tsokas P, Grau-Perales A, Lesburguères E, Bukai J, Khanna K, Chorny J, Chung A, Jou C. Persistent increases of PKM ζ in memory-activated neurons trace LTP maintenance during spatial long-term memory storage. *European Journal of Neuroscience*, 2021, 53(8):6795-6814.
- [20] Liu X, Liu S, Li X, Zhang B, Yue C, Liang S. Intelligent tool wear monitoring based on parallel residual and stacked bidirectional long short-term memory network. *Journal of Manufacturing Systems*, 2021, 60:608-619.
- [21] Kayé B K B, Diaby M, N'Takpé T, Oumtanaga, S. Managing an External Depot in a Production Routing Problem. *IJACSA*, 2020, 11(5): 321-330.
- [22] Adi P D P, Kitagawa A. Performance evaluation of LoRa ES920LR 920 MHz on the development board. *International Journal of Advanced Computer Science and Applications*, 2020, 11(6): 12-19.
- [23] Xie Z, Gu X, Shen Y. A Machine Learning Study of Predicting Mixing and Segregation Behaviors in a Bidisperse Solid-Liquid Fluidized Bed. *Industrial Engineering Chemistry Research*, 2022, 61(24):8551-8565.
- [24] Dandil E, Bier A. Automatic grading of brain tumours using LSTM neural networks on magnetic resonance spectroscopy signals. *IET Image Processing*, 2020, 14(10):1967-1979.
- [25] Tao B, Pejman T. Attention-based LSTM-FCN for earthquake detection and location. *Geophysical Journal International*, 2021, 228(3):1568-1576.
- [26] Jian C, Yang M, Zhang M. Mobile Terminal Gesture Trajectory Recognition Based on Improved LSTM Model. *IET Image Processing*, 2019, 13(11):1914-1921.
- [27] Bao J, Wang X, Zheng Y, Zhang F, Sun P. Lightning Performance Evaluation of Transmission Line Based on Data-Driven Lightning Identification, Tracking, and Analysis. *IEEE Transactions on Electromagnetic Compatibility*, 2020, 63(1):160-171.

Methodological Insights Towards Leveraging Performance in Video Object Tracking and Detection

Divyaprabha¹, Dr. M.Z Kurian²

Research Scholar¹, Professor²

Dept. of Electronics & Communication Engg, Sri Siddhartha Institute of Technology, Tumkur, India^{1,2}

Abstract—Video Object Detection and Tracking (VODT), one of its integral operations of surveillance system in present time, mechanizes a way to identify and track the target object autonomously and seamlessly within its visual field. However, the challenges associated with video feeding are immensely high, and the scene context is out of human control, posing an impediment to a successful model of VODT. The presented work has discussed about effectiveness of existing VODT approaches considering its identified taxonomies viz. satellite based, remote sensing-based, unmanned-based, Real-time Tracking based, behavioral analysis and event detection based, integration of multiple data sources, and privacy and ethics. Further, research trend associated with cumulative publications and evolving methods to realize the frequently used methodologies in VODT. Further, the results of review showcase that there is prominent research gap of manifold attributes that demands to be addressed for improving performance of VODT.

Keywords—Object detection; object tracking; video; visual field; surveillance system; video feed

I. INTRODUCTION

Object detection and tracking are essential operations that any surveillance system demands [1]. Irrespective of archives of research models presented to date, there is still an open concern associated with object detection and tracking [2]-[5]. One of the significant shortcomings of cumulative research work is the lack of any model which can guarantee maximized performance, higher accuracy and dominantly robust [6]-[8]. The prime challenges are encountered in this domain mainly because of the condition stated towards tracking and detecting an object for a given scenario. The complexity of such implementation depends upon the use cases. In the case of objects with fewer visual features, it is subjectively easier to detect all the features associated with that visual characteristic. However, complexity arises for the object of dynamic type, where extraction of features is quite a complex process in the presence of challenging background and foreground situations. Predominant challenges surface regarding Video Object Detection and Tracking (VODT), especially for moving objects in general. However, there could be multiple mobility situations and statics of either foreground or background [9]. Various other factors that impose challenges in VODT are occlusion (full/partial), fluctuation in illumination condition, deformation of the target, and variability in the pose of the target object. Various standard approaches of detection consist of object detection based on i) features (color, shape) [10], ii) template (deformable / fixed) [11], and iii) motion (global energy, statistical test, thresholding) [12]. On the other, the

approaches of tracking are carried out using motion information mainly (region, boundary) [13]. The conventional mechanism of VODT emphasizes two essential attributes, i.e., information related to the motion of the target object and visual features (e.g., shape, texture, color, etc.). Owing to the variable nature of such attributes, it is always better to perform the modeling of VODT by integrating temporal features and statistical models with visual features [14]. Normally, the process of VODT consists of obtaining regions based on visual features using frame segmentation followed by combining all regions characterized by equivalent motion vectors. The majority of the existing approaches of VODT are witnessed with the adoption of multiple forms of approach as well as a combination of approaches. The area of implementation is so scattered that there is a lack of any uniform approach towards VODT with a consistent track of performance. It is essential to offer a comprehensive insight into various study models towards VODT to identify the possibilities of future research work by reviewing its strength and weakness. Therefore, the proposed study discusses existing approaches to offer insight into the effectiveness of existing VODT. The contributions made in the proposed paper are:

- Commercial usage of VODT is discussed to understand global deployment in the practical world,
- Existing approaches are classified concerning some potential use cases of deployment to understand the effectiveness in deployment scenarios,
- Discussion of research trends is carried out to offer insight into identifying frequently used techniques,
- A crisp discussion of the research gap associated with existing VODT techniques.

The organization of this paper is as follows: Section II discusses insight into the commercial application of VODT, followed by an elaborated discussion with the classification of VODT approaches in Section III. A discussion of the most frequently adopted video dataset is carried out in Section IV. At the same time, the research trend is discussed in Section V. Highlights of the results discussion is carried out in Section VI, while conclusive remarks are stated in Section VII.

II. COMMERCIAL USAGE OF VODT

The commercial usage of ODT mainly performs localization, identification, tracking, counting, and recognizing the anomalies of an object of specific form present within the captured image frame of the Video. The ranges of such objects

are quite wide enough. The devices that perform such tasks consist of recognition and classification modules. In general, ODT from the video scene is carried out by extracting the object from the background image, presenting an anticipated equivalent object class proposition. Finally, a bounding box is constructed to encapsulate the object. This section discusses the different use cases of the application of ODT.

A. Smart City Usage

With the advent of the Internet-of-Things (IoT), the environment of a smart city is now equipped with multi-functional sensors and sophisticated devices for monitoring. Various ODT applications find a suitably higher scope in smart cities, viz.

- **Monitoring of In-Cabin Space:** Modern technologies are now used for monitoring the environment inside a closed space like a home or vehicle to understand people's behavior. This is done using computer vision, eye tracking, pose estimation, etc. It is specifically helpful in detecting drivers' drowsiness to make the ride safer [15].
- **Occupancy of Parking Site:** The mechanism of VODT in computer vision is used for classifying vacant and occupied parking areas. The information can be transmitted in real-time directly to the driver directing them towards the target open end parking space [16].
- **Counting People:** This application tracks visitors' counts to plan security in public areas. They are usually deployed in transit areas, which can capture data on incoming and outgoing people for security and capacity management [17].
- **Monitoring Traffic:** Various Road and traffic conditions can be monitored via VODT. Multiple zones in the city can be monitored to identify congestion or violation in driving or accidents. It is also used for identifying the vehicle's registration details by capturing the license plate image [18].
- **Autonomous Driving:** Unmanned driving vehicle is the next vehicle level with the advancement of computer vision technology. Such application uses VODT to identify traffic signs, markers on the lane, vehicles, pedestrians, etc. [19].

B. Industrial Usage

ODT is also used for various industrial applications on a large and small scale. The following are the different uses of VODT:

- **Enhancing Productivity:** Multiple activities of the workers in different locations, e.g., construction sites, production facilities, and warehouses, can be monitored using VODT. The worker's activities can be monitored while retaining private information [20].
- **Detection of Defect & Anomaly:** Different forms of industrial products can be assessed using computer

vision to identify the defects or sub-standard quality in finished products. Various processes associated with quality control, production lines, and workstations can be monitored using VODT [21].

- **Product Assembly:** Adoption of ODT can be used for ensuring the selection of appropriate components over the assembly lines. Different forms of automated production systems and robotics are facilitated with various intelligent feeds from the ODT system [22].
- **Detection of PPE:** ODT can also be used to ensure worker safety. This is done by evaluating if the employees have put on a specific form of Personal Protective Equipment (PPE) assigned by the safety standards. The system can distinguish between employees with and without protection gears [23].

C. Retail Usage

The usage of computer vision plays an important role in the retail industry with multiple purposes. The retail sector has manifold concerns about reviewing the user's response, behavior, sales processes, etc. Various retail-based applications are:

- **Customer Experience:** ODT can extract the customer's actions during store visits. It can assist in understanding the possibilities of assistance that any visiting customer may require [24].
- **Analysis of Foot Traffic:** This is nearly similar to object counting applications under smart city usage. This application is used for counting the incoming and outgoing visitors in the store to understand the peak traffic of visitors. The manager can directly analyze this traffic information, assisting in product placement or promoting the product at a specific location [25].
- **Inventory Management:** This application is used for evaluating the availability of products on specific shelves or warehouses. In case of a drop in product availability, the ODT-based system can forward the notification to the inventory manager to restock. Artificial intelligence effectively controls inventory management systems [26].
- **Contactless Checkout:** Contactless kiosk is used for a contactless payment system, an effective solution for crowd and queue management inside the store [27].
- **Video Analytics:** ODT can be used for capturing the feed inside retail to evaluate customers' shopping behavior. Potential analytical information can be obtained from such form of Video feeds powered by artificial intelligence. It can also assist in faster and more effective service delivery within the store with less customer wait time [28].

Therefore, all the taxonomies mentioned above of application for VODT are some of the prime focuses on commercial and research interests. The next section discusses the current approaches of VODT.

III. CURRENT RELATED WORK OF VODT

At present, there are various categories of approaches as well as methodologies that have been undertaken towards investigating VODT problems. However, categorization concerning methodologies is quite a tedious process as it is noted that sometimes the same methodology is adopted to solve two different use cases of problems. Hence, this part of the discussion is based on various use cases of problems being investigated and highlights individual methodologies used to address the problem.

A. Satellite based VODT Approaches

This mechanism calls for capturing the satellite's video feed and applying algorithms to identify the objects from the feeds. It is one of the most evolving and challenging investigation trends in VODT, irrespective of progressive research [29]. Most of the existing research approach targets solving the identification and tracking of smaller objects from satellite video datasets. The prime challenge is to identify the foreground objects, which are smaller and have low contrast. The work carried out by Chen et al. [30] has developed a scheme for enhancing the tracking performance of an object of smaller sizes captured from satellite-annotated images. Hu et al. [31] have constructed a network based on a regression model integrated with gradient descent and convolution layers. The tracker is designed from the background context using a regression network. A deep neural network is used to train motion and appearance features. Shi et al. [32] have implemented a technique for detecting and tracking mobile aircraft. The investigation carried out by Wu et al. [33] constructed an enhanced filter with kernel correlation to track the smaller-sized object. According to this experiment, the mean peak response is integrated with the mean energy of peak correlation associated with the response map to mitigate occluded objects.

Further, an adaptive Kalman filter is used to improve the tracking performance. Xuan et al. also uses a similar methodology line [34] where the motion trajectory is integrated with the Kalman filter to track the smaller object. The study outcome is claimed of 95% accuracy in tracking performance. The work by Zhang et al. [35] has integrated features of optical flow with a Histogram of oriented Gradient (HoG) with a target towards enhancing the target representation. The boundary effects are mitigated by integrating the inertial mechanism and Kalman filter, while the interference attenuation is accomplished using the disruptor-aware process. Zhou et al. [36] performed a unique study considering satellite images where a pyramid network of selected features is presented to address the problem of computing gradient inconsistencies. The study has also introduced a contrastive learning mechanism to represent an object robustly. The work by Zhu et al. [37] used Siamese deep network designed to improve the smaller object representation from satellite videos. This model aims to extract features from the search and template branch in the Siamese network while the target position is determined from the search branch. The model's outcome is proven to offer a satisfactory accuracy evaluated over multiple performance parameters.

B. Remote Sensing-based VODT Approaches

This mechanism identifies and evaluates the physical characteristic associated with a region by computing the radiation emitted from a specific distance. Usually, such monitoring is carried out from aircraft or satellites using specific feed-capturing devices. The present research is being carried out towards remote sensing-based VODT using a conventional mechanism [38] that consists of simplified object detection and tracking. Detection is facilitated by various techniques, e.g., background subtraction process, optical flow mechanism, and method of computing frame difference. The work carried out by Lei and Guo [39] implemented a technique that can detect and track multiple objects considering the Gaussian mixture model for extracting road networks using deep learning.

Further, a neighborhood search mechanism is implemented for tracking connected with the data association method. The neural network adoption is witnessed in Lin's [40] work, where tracking is carried out for multiple targets. The implementation is carried out considering the integration of real-time tracking of the target using deep learning, combined geometric features, and a dual neural network. The study model also constructs an optimization mechanism using the least squares method, where the constrained residual terms are developed over the pixel plane. A unit for inertial measurement is developed. Finally, a mathematical model is constructed for multitarget tracking. The mechanism presented by Ma [41] discusses enhancing the probability of identifying objects with poor signal quality and minimizing the number of outliers. Another unique study has been presented by Tochon et al. [42] on chemical gas plume tracking. Considering hyperspectral Video sequences, the author has used object detection sequentially, considering temporal, spatial, and spectral information. The work carried out by Uzkent et al. [43] [44], where a fusion mechanism of kinematic likelihood is estimated for analyzing hyperspectral information. The idea is to detect and track mobile objects captured from aerial Video over a short time window.

Further, the author has also used deep features for improving the tracking performance using a convolution neural network and kernelized correlation filter. However, the study model cannot process an isometric view of the scene. This problem is reported to be addressed in the work of Wei et al. [45], where a three-dimensional view is considered with total variation towards tracking a moving object.

C. Unmanned VODT Approaches

Various existing VODT use cases have considered the Video captured from Unmanned Aerial Vehicle (UAV) system. The work carried out by Cintas et al. [46] has constructed a model that can perform tracking of UAV using vision factor from another UAV using a deep learning approach and kernelized correlation filter. Deng et al. [47] used a regularization method using spatial and temporal attributes and constructing an enhanced filter for discriminative correlation. The technique also uses a joint optimization mechanism to improve target representation and reduce distractors. A unique method is implemented by Ding et al. [48], where object detection for UAVs is carried out by blockchain and hash-based approach. According to this implementation, a hash-

based network is constructed for extracting the hash representation of an object where further recovery of tracking interrupts and feature fusion is carried out. The study model also integrates motion features with deep hash features to address the problem of object occlusion. Detection of a small object is carried out by Liang et al. [49] using a fusion of features and Detection of a single shot based on scaling operation. A feature pyramid is constructed using average pooling operation. The fusion of the feature module and the deconvolution module is used for generating the feature pyramid. Lin et al. [50] have designed a correlation filter based on blocks' bidirectional incongruity towards tracking objects. Buhler et al. [51] have carried out a different form of work that doesn't work on video object detection or tracking algorithm but offers a robust platform to perform such tasks. The authors have used remote sensing images of snowflakes to classify snow types' normal to critical states. Shan et al. [52] have developed a ship-tracking mechanism considering maritime data over multiple videos with manual annotation. Xue et al. [53] have developed a semantic-based framework for object tracking to improve the performance of correlation filters of discriminative nature. The study model generates the specific region of interest followed by filtering where the semantic coefficient prediction is carried out. The study model also contributes towards reducing parametric redundancy by sharing the object's semantic segmentation information and network layer. The adoption of a correlation filter is also witnessed in the work of Ye et al. [54]. The study model applies a concept of multi-regularization toward constructing a correlation filter to facilitate the tracking and localization of an object. The work carried out by Yu et al. [55] has developed a region of interest while object detection is carried out from radar and photoelectric cell.

D. Miscellaneous Approaches

Various methodologies are being introduced, considering different forms of use cases. The work carried out by Banerjee et al. [56] has developed a hardware design of architecture for the acquisition of Video. The study model uses dynamic programming towards the reduction of normalized area for the allocation of resources. Further, the Kalman filter is used for performing tracks. A simplified mechanism of object detection and tracking is carried out by Chen et al. [57], where a region of interest in a closed loop is developed to magnify the field of view, considering spatial and temporal attributes of the feed. Cheng et al. [58] have developed a framework to track commodities where the edge computation caters to resource dependencies. The model can also track multiple videos while significantly controlling computation overhead.

Further, the Markov model is designed for compensating the missing data. A similar form of edge computing investigation is also carried out by Gu et al. [59], where a Siamese convolution network is constructed to carry out object tracking. A collaborative architecture of cloud and edge computation is designed for this purpose. The adoption of the Siamese network is also reported in work carried out by Lee et al. [60] towards object tracking. According to this model, a bounding box encapsulates the object, followed by tracking using a Siamese network which minimizes the computational complexity. Object detection is carried out towards each

classified object in Video, followed by performing tracking using sequential frames. All the obtained retargeted frames are obtained by rearranging all frames.

A discreet and unique model of object tracking is designed by Liu et al. [61], which addresses the problem of the extreme degree of occlusion. It also assists in the classification of similar forms of objects. The study model uses a depth tracking mechanism and presents a matching strategy of two look-alike objects using semantics using indoor scene video. The study by Marshall et al. [62] developed a three-dimensional tracking mechanism of an object of radiological or nuclear type. The study uses a Kalman filter to perform tracking using multiple cameras of a specific form. Mostafa et al. [63] have developed a computational model of multi-object tracking using the Kalman filter for tracking crowds. According to this model, a unique encoder model is designed to generate computationally efficient features.

Further, a linear transformation is carried out to retain maximum accuracy. Ramesh et al. [64] have presented a framework for long-term tracking over dynamic motion. Considering a usual tracking condition, the proposed scheme is implemented considering a moving camera where a local sliding window is formulated to confirm the higher reliability. The analysis considers quantitative analysis, rotation, and translation in a laboratory environment. The work carried out by Ren et al. [65] has discussed a computational model that can perform both tracks as well as counting from the crowd scene using a network flow approach. Sun et al. [66] have implemented a study model for tracking multiple objects. The presented study uses deep learning to optimize the representation of an object and their respective affinities occurring over every frame. An affinity-based deep learning network is designed to track objects present and disappearing from the frames. Eventually, the problem of the missing object is addressed in this work.

E. Robust Real-Time Object Tracking

The practical applications of object tracking demand its operation of real-time feed of video stream which is one the most challenging task. At present, there are certain research work which has been dedicated purely towards ensuring robust real-time object tracking. The work carried out by Zhang and Ren [67] have used Kalman Filter and Kernel Correlation Filter in order to carry out object detection. Further, missed detection is addressed by updating the tracking box as per the rate of change in scaling. Backstepping is used for catering the design of kinematic controller with respect to Lyapunov stability. Further, the work of Cao et al. [68] have addressed the issues associated with tracking drift and loss of object by using Siamese network with double template while the feature extraction is performed by enhanced MobileNet V2. Similar work is also carried out by Zhao et al. [69] where Siamese Network is used for estimation of optical flow based on feature pyramid approach. The movement attributes are further evaluated using mapping of pyramid correlation between two contiguous frames. The study also addresses the problems associated with ambient noise using channel and spatial attention. Another study of real-time object tracking is carried out by Du et al. [70] that addresses the problems associated with implementing Correlation Filter-based Trackers (CFT) on

practical grounds. In this mechanism, feature extraction is carried out using Histogram of oriented Gradient (HoG) followed by integration of scale adaptive, target re-detection, and discriminative appearance model. Although, there are various research work carried out towards object tracking, but only few of the recent work has been claimed towards real-time tracking. Another perspective is that these studies doesn't address overall agenda-based attributes that is demanded for real-time tracking e.g., initialization and re-detection, computational constraint, fast motion, motion blur, perspective and scale changes, and occlusion all together. Only few of the above-mentioned attributes have been chosen in current works on real-time object tracking.

F. Behavioral Analysis and Event Detection

There are various unique studies being carried out towards object detection in the perspective of event detection and behavioral analysis. The work of Chakole et al. [71] has addressed the issues pertaining to anomaly detection for the use-case of crowd behavior using correlation-based optical flow. Object detection has also been studied with respect to recognition mechanism of human activity. A typical structure of recognition of human activity is designed by Vrskova et al. [72] using deep learning approach. The prime motive of this work is to address the challenge associated with the less availability of such dataset, where the authors have constructed a new dataset consisting of abnormal activities. Study associated with anomaly detection is carried out by Chang et al. [73], Yahaya et al. [74], and Yang et al. [75]. Such methodologies are meant to identify a discrete vector of motion, which is further subjected to varied strategies in order to perform behavioral analysis. Irrespective of varied approaches used for such form of behavioral analysis, they still have limiting attributes associated with their practical utilization on real-time perspective.

G. Integration of Multiple Data Sources

Majority of the conventional implementation scenario calls for implementing its detection and tracking algorithm considering single sources of data. However, there are certain research work where such objectives are accomplished using multiple sources of data. One such work has been reported by Rehman et al. [76] where visual and audio signals were used as input stream for detection of anomaly. According to this study, the model has integrated particle swarm optimization with optical flow for obtaining visual features. The acoustic features have been obtained from volume, rate of zero crossing, energy, spectral flux etc. Similar pattern of methodology is also created by Benegui and Ionescu [77] which carry out authentication using feeds from camera and motion sensors. Deep neural networks are used using Support Vector Machine for generating an embedding vector obtained from transformed signals. Another similar form of study approach is seen in work of Shin et al. [78] where heterogeneous sources of data e.g., temperature, elevation changes, pattern-specific behavior of human have been considered. The prime notion of this study is to improve the accuracy towards detection of anomaly associated with surveillance map. Majority of the above-

mentioned studies have been carried out by integrating multiple and different sources of data in order to obtain an embedded vector which are further subjected towards analysis towards detection and tracking.

H. Privacy and Ethics

Privacy is one of the essential attributes to be protected while performing object detection. The work carried out by Dave et al. [79] has developed a scheme towards privacy preservation while performing recognition of human action. The authors have used supervised schemes towards removing the private details without any dependency of labelling. The work carried out by He et al. [80] have implemented a mechanism of object blurring as well as object swapping in order to preserve private information while performing object detection. Zhang et al. [81] have used federated learning scheme along with blockchain for detection visual object along with retention of private information. This scheme uses encryption as well as validation nodes in order to resist any form of privacy-related attacks. Further mechanism of object detection along with privacy preservation is carried out by Bai et al. [82] where Convolution Neural Network (CNN) has been utilized along with secret sharing scheme towards protecting private information captured from vehicular edge computing. The prime motive of such approaches is basically to adopted varied privacy preservation approaches ensuring that it won't hamper any parameters potentially responsible for detection of an object.

Hence, it can be seen that there are various ranges of methodologies being developed in current times on various perspective of object detection. It is eventual that all these approaches leverage the performance of object detection addressing issues on local levels and use case.

Apart from this, various works is carried out in conventional VODT with different techniques. One common fact observed in all the miscellaneous techniques is that tracking is more emphasized than preliminary detection operation. As tracking performance completely depends upon the logic configured for the detection module, the performance evaluation doesn't highlight this fact. Apart from this, all the usual and generalized VODT schemes discussed in this section can be used for multiple purposes and assessed over different datasets; however, their applicability towards functional over complex scenarios or challenging video environments is reportedly not investigated. Another significant observation is that most schemes utilize varied principles of framing up the dimension of an object to track the objects in the video stream. However, there is no report of any work towards the generalized object tracking model utilized in the video stream to ensure better analytical operation toward improving the tracking performance. Further, it is noted that there is different use-case specific implementation study where generalization is challenging to be achieved. Table I highlights the summarized version of the discussion in this section concerning methods used for addressing identified research problems, advantages, and limitations.

TABLE I. SUMMARY OF RELATED WORK IN VODT

Author	Problem	Methodology	Advantage	Limitation
Chen et al. [30]	Tracking of the smaller object	Historical model, Kalman filter	Higher Accuracy	Lacks scale adaptive process
Hu et al. [31]	Drift in tracking	Convolution Neural Network, Regression	Better tracking performance	Orientation and position encoding are not feasible
Shi et al. [32]	Tracking mobile aircraft	Shift invariant feature transform	The better predictive calculation for flight path	Not applicable for multiple object tracking
Wu et al. [33]	Tracking of the smaller object	Adaptive Kalman filter	Mitigates occlusion problem	The iterative process leads to a computational burden
Xuan et al. [34]	Tracking of the smaller object	Kalman filter, averaging motion trajectory	95% of accuracy, effective processing speed	Restricted to single object tracking
Zhang et al. [35]	Tracking of the smaller object	Kalman filter, HoG, inertial mechanism	Can perform multi-object tracking, effective classification	Doesn't overcome occlusion problem
Zhou et al. [36]	Gradient inconsistencies	Pyramid network with selected feature scale	Effective reduction of interface	Doesn't consider temporal factors while tracking
Zhu et al. [37]	Inferior distinguishability of objects	Siamese deep network	Offers real-time processing capabilities	It doesn't address the impact of padding and shallow network
Lei and Guo [39]	Detection and tracking	Gaussian Mixture Model, Neighborhood Search	A simplified approach supports multi-object Detection	No benchmarking
Lin [40]	Multitarget tracking	Deep learning, dual neural network	Retain superior tracking quality	Narrowed extensive analysis
Ma [41]	Addressing Outliers in Detection	Probability-based modeling to reduce outliers	Improved rate of Detection	Lower scope of analysis
Tochon et al. [42]	Tracking of chemical gas	Sequential object tracking using temporal, spatial, and spectral information	Reduced computational time	Specific to the composition of chemical gas
Uzkent et al. [43][44]	Tracking of vehicle	Fusion mechanism for kinematic likelihood, convolution neural network, and kernelized correlation filter.	Independent of manual allocation of the threshold for multiple objects	No scope for an isometric view of the scene.
Wei et al. [45]	Object tracking	Principal component analysis	Simplified mechanism of implementation	No benchmarking
Cintas et al. [46]	Tracking UAV from another UAV	Kernelized correlation filter, neural network	82.7% of accuracy, developed a new dataset	Demands more real-time testing to confirm the outcome applicability
Deng et al. [47]	Limited computational capability in UAV tracking	Regularization, correlation filter, joint optimization	Better tracking performance	Higher processing time
Ding et al. [48]	Multiple object tracking	Blockchain, deep hash	Mitigates object occlusion	Consumed higher resources
Liang et al. [49]	Detection of a small object	Spatial context analysis	Higher accuracy	Demands higher resources to map with the outcome
Lin et al. [50]	Improving correlation filter	Discriminative correlation filter	Support real-time tracking with low resources	Outcome specific to data
Buhler et al. [51]	Snow type classification	The reflective property of snow-based modeling	Effective classification strategy	Study model specific to the object
Shan et al. [52]	Ship tracking	Multiple maritime dataset evaluation	Comprehensive analytical approach	Study model specific to the object
Xue et al. [53]	Correlation tracking	Semantic segmentation	Minimize parametric redundancy	It doesn't address the occlusion problem effectively
Ye et al. [54]	Improving correlation filter performance	Multi-regularized filter development	Higher accuracy	Not analyzed on broader assessment environment
Yu et al. [55]	Object tracking	Prediction based on the region of interest, object extraction from the first frame	Reduced processing time	Tracking performance degrades over dynamic background
Banerjee et al. [56]	Acquisition of object	Viterbi, Kalman filter quadtree segmented Video, convolution neural network	Extensive test cases	Needs a broader extensive assessment environment
Chen et al. [57]	Object detection & tracking	Region on interest on a closed loop, dual resolution	Higher resolution	The computationally extensive mechanism for identifying a region of interest
Cheng et al. [58]	Commodity tracking	Edge computing, Markov model	Reduces computational burden	It needs more extensive analysis with recent models
Gu et al. [59]	Object tracking	Siamese Convolution Network, cloud, and edge environment	Reduced energy consumption, Higher accuracy	No extensive analysis
Lee et al. [60]	Object detection	Siamese Convolution network	Higher similarity score	Highly iterative scheme
Liu et al. [61]	Multiple indoor object tracking	Semantic strategy to match objects	Simplified and efficient tracking	It depends upon the illumination condition
Marshall et al. [62]	Object detection	Kalman filter, neural network, non-negative matrix factorization	Reduces anomaly count	Applicability is Specific to this use case only

Mostafa et al. [63]	Crowd tracking	Encoder, Kalman filter	Control over computation time	It depends upon the illumination condition
Ramesh et al. [64]	Object tracking	Bayesian Bootstrapping, sliding window detector	Faster Detection	Assessed over laboratory confined setting.
Ren et al. [65]	Object tracking, Detection, counting	Network flow programming	Supports Multiple operations	The optimization of the counting problem is not addressed
Sun et al. [66]	Tracking multiple objects	Affinity-based deep learning,	Effective analysis	Induces computational complexity over the long run
Zhang and Ren [67]	Real-time object tracking for robots	Kalman Filter, Kernel Correlation Filter	Higher stability of target tracking	Doesn't assess model uncertainty
Cao et al. [68]	Tracking drift, object loss, inadequate robustness	Siamese Network (double template), MobileNet V2	Higher tracking accuracy	Consumes more training efforts
Zhao et al. [69]	Interference of cluttered background	Siamese Network, Pyramid Correlation Mapping, Channel/Spatial attention	Higher success rate of tracking	Time consuming classification
Du et al. [70]	Issues in CFT in real time	HoG, discriminative adaptive model, target re-detection	Simplified architectural implementation	Cannot be applied for objects with consist orientation changes
Chakole et al. [71]	Anomaly detection of crowd	Correlation of optical flow	Satisfactory detection	Not benchmarked
Vrskova et al. [72]	Recognition of abnormal human activity	Long Short-Term Memory	96% of classification accuracy	Induces higher computational burden
Chang et al. [73], Yahaya et al. [74], Yang et al. [75]	Anomaly detection	k-means, autoencoder, support vector machine, deep learning	Multi-target anomaly detection	Some of the outliers are also subjected to training
Rehman et al. [76]	Anomaly detection	Multi-modal (visual and acoustic source)	Improved accuracy.	Restricted to limited data, no extensive analysis over broader dataset size
Benegui and Ionescu [77]	Authentication using difference sources	Deep Neural Network, Support Vector Machine	Higher accuracy score	Not much suitable for larger dataset
Shin et al. [78]	Representation of multimodal data for anomaly detection	Analytical model using temperature, elevation changes, pattern-specific behavior of human	Ideal for surveillance map in smart city	Study not benchmarked
Dave et al. [79]	Privacy preservation during recognition of human activity	Self-supervised learning	Non-dependency of labelling	Model restricted to specific use-case only
He et al. [80]	Privacy preservation	Object swapping and blurring	Retains better accuracy and offers imperceptibility	Not proven using complex forms of object
Zhang et al. [81]	Privacy preservation from visual object detection	Blockchain and federated learning	Offers potential security from maximal threats	Higher cost of implementation
Bai et al. [82]	Privacy preservation	Multiple secret shares, CNN	Offers multi-stage detection, and optimal privacy	Higher resource dependencies

IV. DATASET AND PERFORMANCE METRIC

After reviewing various approaches and research models, the dataset reportedly uses different forms. The frequently used dataset is *ImageNet VID* [83]. It is a benchmarked dataset with training videos of 3862 and validation videos of 555. The frame rates are maintained at 20-30 frames per second while an annotation is provided on all the Videos. Another commonly used dataset is EPIC KITCHEN, which has predefined 290 classes of an object with bounding boxes present in video samples of 32 kitchens [84]. There are also standard datasets used for specific purposes in VODT, viz. i) generalized object detection from the PASCAL dataset [85][86], ii) pedestrian data from Caltech [87], iii) indexing and retrieval of Video from TRECVID [88], iv) categorization data of human activities from HMDB-51 [89], v) annotated and segmented data of sports from Sports-1M [90], vi) tracking of an object from the MOT dataset [91] and VOT dataset [92], vii) detection of a mobile object from the CDnet2014 dataset [93], and viii) segmentation of an object from the DAVIS dataset [94]. Irrespective of availability of wider ranges of dataset, there is considerably a smaller number of research implementation towards involving multi-modal approaches of VODT.

Various performance metrics are deployed while assessing the effectiveness of VODT; however, they are all connected with accuracy-based attributes. The most common performance metric is *mean Average Precision* (mAP), used for assessing an object's traditional detection and tracking. The metric mAP is associated with accuracies of classification and regression. Before evaluating using mAP, the existing system also evaluates confidence scores and true positives. Based on the speed of the mobile object, the inference of mAP is specified in the form of fast, medium, and slow. Certain discussion states that mAP cannot be solely used for performance evaluation as it cannot capture its temporal attributes [95]. Hence, another metric termed *average delay* has surfaced, which computes the number of video frames considered for Detection and tracking from the initial frame. A dataset named ImageNet VIDT was used to verify the appropriateness of the average delay. The study outcome states that satisfactory average precision can be confirmed if the method is witnessed to maximize the average delay value. On the contrary, the maximized value of average delay will also signify maximized delay in Detection and tracking performance. Therefore, it can be stated that if any method deploys average precision as a sole performance metric, then it is challenging to evaluate the actual average delay score. It will eventually conclude that average precision

is insufficient to represent temporal attributes of the assessing framework for VODT. Hence, it is suggested to fine-tune the performance metric based on the research problem undertaken in the study model. The next section discusses the analysis of the research trend.

V. RESEARCH TREND ANALYSIS

From the prior sections, different methodology variants are being deployed toward VODT. However, it is also noted that various common methodologies are being used to consider different use cases of VODT. It is noted that use cases play a critical role in classifying VODT methodology owing to the inclusion of discreet challenges and characteristics. Hence, to understand research trends, the primary assessment has been carried out to identify the popularity of consideration of such use cases in research publications. Fig. 1 highlights the trends of publication of such use-case adoption in VODT.

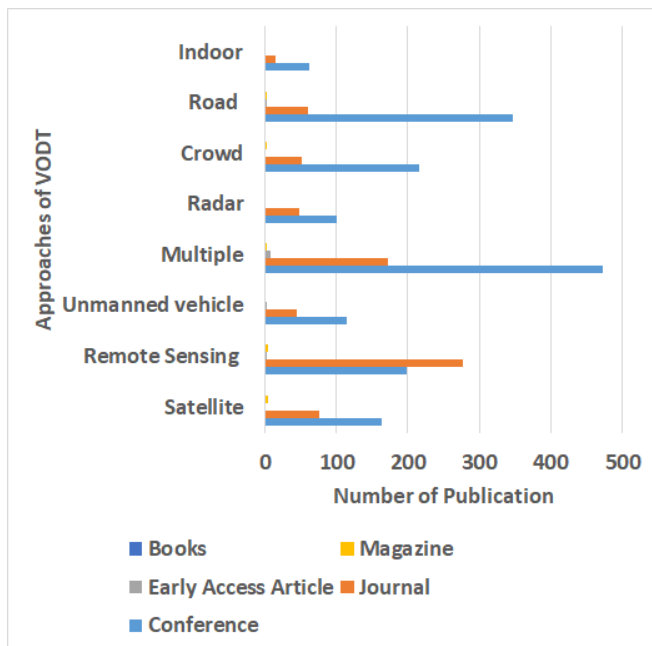


Fig. 1. Trends towards research publication.

The observation towards arriving at the graphical outcome in Fig. 1 is carried out from several different research papers published between 2017 to 2022 from IEEE Xplore, Springer, Elsevier, Wiley, ACM, MDPI, etc. The total number of count of papers for the allocated use case is more than the exhibited figure. This is because the data for Fig. 1 is obtained by filtering out only the significant experiments, excluding all the discussion or conceptual-based publications. The idea was to grab the information associated with including prominent methodologies involved in solving challenges of use cases in VODT. From this graphical outcome of the trend, it is noticed that a greater number of studies are concentrated on remote Sensing and multiple object detection method. Nearly an equivalent number of studies are on satellite-based approaches, unmanned vehicle-based VODT, radar-based VODT, and road/crowd-based VODT. Studies towards indoor-based VODT are significantly fewer. The analysis is also carried out to identify trends in methodologies by removing the constraint of year-based publication. It is noted that evolving trends for

VODT are flow-based, LSTM-based, attention-network-based, tracking-based, and miscellaneous approaches, as stated in Fig. 2.

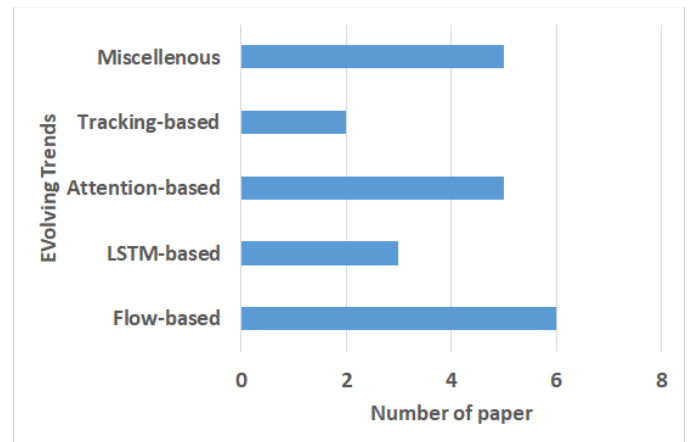


Fig. 2. Observation of evolving trend.

A closer look into multiple evolving strategies in Fig. 2 concludes that the VODT mechanism is broadly classified concerning feature aggregation and propagation. *Flow-based mechanism* adopts optical flows in dual directions. Significant contribution work was noted towards adopting this technique, viz. Deep Feature Flow [96], feature aggregation with guided flow [97], impression network [98][99][100], flow algorithm using the difference between adjacent frames and adoption of time and spatial factor of the frame, deep learning over FlowNet [101]. Further, adopting *Short-Term Long Memory (LSTM)* is another evolving trend toward maximum utilization of both time and spatial factors associated with video frames. Some of the significant contributions have been carried out by introducing unique techniques, e.g., convolution LSTM [102], online [103], and offline LSTM methods [104] [105]. *The attention-based technique* can analyze the long-duration Video for aligning the feature map with a target to minimize dependency on computational resources. The implication of such mechanism was noted in the form of various approaches viz. approaches considering only temporal factors locally [106] and global factors [107], hybrid method integrating both local and global time-based factors [108], regression and classification-based fusion of feature maps [109], managing external memory with guided objected for global aggregation [107]. The next technique is *tracking-based*, where temporal information is utilized for object detection over video frames with a fixed interval. Some of the significant approaches noted for this approach are adaptive frame-based tracking [110][111], construction of forward and backward trackers [112], adoption of refinement network for integrated Detection and tracking [113], and convolution network-based tracking [114]. Apart from the above-mentioned evolving trends of approaches, there are also *miscellaneous* methods [56]-[66]. The analysis of this approach is further carried out from the perspective of time-based evolution. Fig. 3 highlights the year-wise proposition of the above-stated approaches. The adoption of the ImageNet VID dataset [83] was the first to be evolved in 2015, followed by the evolution of tracking and Detection with cooperation TDC [112] in 2016. There has been progressively developed in the year 2017, which witnessed the Flow of Deep feature FDF

[115] for the first time along with feature aggregation with guided flow FAGF [107] and impression network [100].

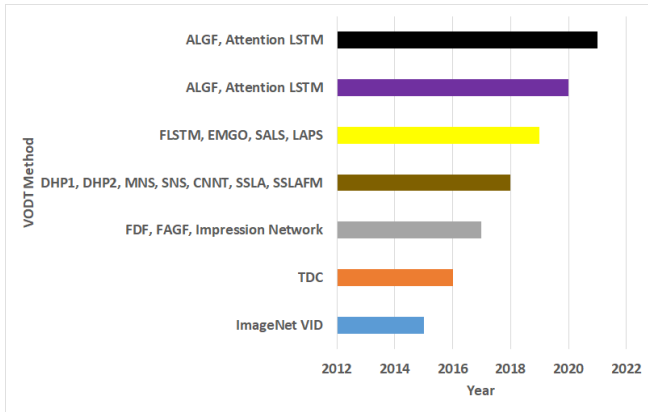


Fig. 3. Occurrence of evolving trends year wise.

The speed of new techniques further multiplied in 2018, where multiple techniques have been reported, viz. Detection with high-performance DHP1 [99] and DHPP2 [98] sampling network with spatiotemporal SNS [116], Memory network with spatiotemporal MNS [117], Convolution Neural Network based Tubelets CNNT [118], Detection of a single shot with LSTM and attention SSLA [119], Detection of a single shot with LSTM and attention with feature map LSTM-SSLAFM [120], etc. A similar pace of research trend was reported in the year 2019, where flow and LSTM have potentially improved along with new techniques of Relation Distillation Network FLSTM [106], external memory with guided object EMGO [107], semantic aggregation of level sequence SALS [109], and local attention with progressive sparsity LAPS [121]. The year 2020 and 2021 has witnessed growth in the aggregation of local and global features with improved memory ALGF [108] along with attention LSTM.

Fig. 4 showcases the effectiveness of evolving trends concerning precision scores considering post-precision (PP) and without post-precision. The outcome shows that the involvement of PP always increases the precision, which directly affects the accuracy performance. The score is found to be good for tracking-based and attention-based approaches.

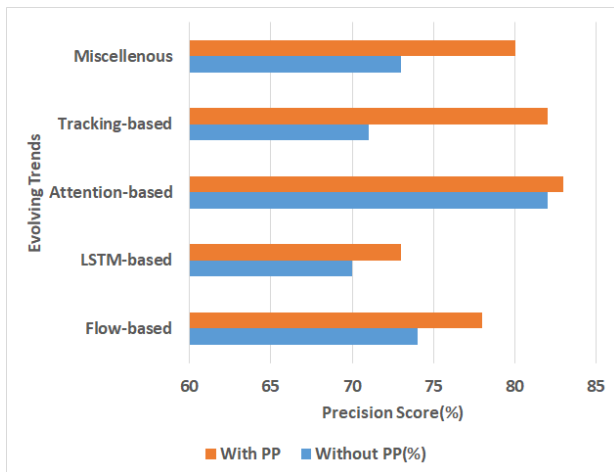


Fig. 4. Precision score for evolving trend.

VI. RESULT AND DISCUSSION

This section presents discussion about gist of learning inference from current review work as well as it also presents a vivid discussion associated with open-end issues of considered methodologies in VODT.

A. Discussion of Current Review

From the prior section, it is noted that there is varied implementation scheme presented towards VODT with unique agenda. One thing is quite clear that every individual research approaches deals with unique set of problems and distinct use-cases. On the other hand, proposed scheme presents review discussion about existing methodologies with an agenda to find its strength and weakness. Hence, it is infeasible to do comparison between implementation-based work and review-based work. However, the study hypothesizes that every research implementation paper must offer a clear-cut highlight of their applicability, about their strength, and about their weakness. Hence, the proposed review considers this common factor to do comparison in order to understand the degree of information associated with merits and demerits linked with suggestion towards further improving VODT. The comparison of the proposed study shown in Table II with the existing considered papers are as follows:

TABLE II. COMPARING PROPOSED SYSTEM WITH CONSIDERED APPROACHES

Approaches	Merit	Demerit
Satellite-based VODT [29]-[37]	Emphasize mechanism to increase accuracy	Doesn't present plan towards increasing adaptivity
Remote Sensing based VODT [38]-[45]	Presents technique to improve detection rate	No discussion towards extensive scope of analysis
Unmanned VODT [46]-[55]	Discusses method to increases accuracy	No solution for increased response time
Real-Time object tracking [67]-[70]	Presents higher stability modelling	No highlights towards addressing less consistent outcomes
Behavioural Analysis/Event Detection [71]-[75]	Supportive of Multi-target detection	No discussion towards increased computational complexity
Data source integration [76]-[78]	Targets higher accuracy	Lack of benchmarking
Privacy & Ethics [79]-[82]	Targets higher security	Higher cost of implementation
Proposed Study	Showcase indicators to balance accuracy and computational complexity	Overlooks security improvements

From Table II, it can be seen that proposed review manuscript offers more information associated with the applicability of various methodologies as well as highlights of various identified issues (discussed in next sub-section) which contributes towards balancing the accuracy demands with computational efficiency demands that is found majorly lacking in maximum considered methodologies in VODT. Further position of current review work with the existing review work is showcased in Table III.

TABLE III. POSITION OF CURRENT REVIEW WITH EXISTING REVIEW

Approaches	Merit	Demerit
Kaur & Singh [3]	Highlight of essential approaches	Emphasize only on deep learning approach
Salari et al. [8]	Informative contents towards scope and challenges in VODT	Lacks highlights of exhaustive research gap discussion
Tulbure et al. [21]	Comprehensive discussion about approaches	Emphasize only on deep CNN approaches
Zhang et al. [29]	Presents methods to deal with complex image-based approaches	Discussion restricted to satellite videos
Proposed Study	Captures all taxonomies of VODT, presents clear highlights of trends, clear cut representation of current research gap	Inference carried out for recent implementation study only

B. Research Gap Analysis

After reviewing the existing schemes of VODT, it is noted that many open-end challenges are still associated with the context of methodologies and use cases considered. A closer look into the efficiencies of all the model exhibit that there is still an open-end issue mainly associated with the speed and accuracy factors balancing with the computational complexities and resource demands. These are some of the areas that have been less emphasized upon. Apart from this, there are few benchmarked datasets which consists of each framed to be labelled. Further, it is noted that ImageNet VID, which is one of the frequently used datasets in VODT, doesn't possess the practical complexities of real scene. Further, the dataset consists of only few objects which are not recommended to be investigated for complex VODT application design. Further, existing schemes either considers global or local information associated with temporal attributes separately. This section outlines the prominent issues being identified in the form of a research gap:

- Issues with the VODT dataset: There are many available datasets for performing VODT, but they lack major benchmarking. The available video dataset doesn't have challenges like a practical environment. Apart from this, very limited objects are present in each frame. Hence, accuracy will always be anticipated to be better while using such video datasets; however, they are less likely to apply to real-world applications.
- Issues with performance metrics: There is no specific standard performance metric for VODT analysis. It should be noted that the performance metric of mAP is extracted from the experiments from object detection of an image. However, mAP cannot process temporal attributes, leading to the evolution of average delay with this evaluation capability. However, this metric cannot evaluate the stability factor associated with the video dataset. Hence, varied ranges of accuracy-based metrics are in practice for assessing the effectiveness of VODT.
- Narrowed Scope of Optimization: An algorithm must encapsulate the maximum constraint condition to

optimize detection and track performance accuracy. A better form of constraint modeling can harness maximum information from the context of Video in the dataset and its associated dynamicity. It demands the utilization of local and global information; however, very few reported studies have such consideration. This imposes an extensive challenge to implement the optimization of VODT.

- Adoption of Complex Method: It is noted that learning-based algorithms are extensively used to improve the overall performance of VODT in many cases. However, such learning algorithms are associated with higher ranges of complexities that are not found to be addressed in existing studies. Adopting such a method (e.g., Convolution Neural Network, Siamese Network) has higher dependencies on training data, indirectly increasing the dependencies of resources required to process it. As a result, it also results in slower computation speed.
- The trade-off between research and practical demand: Although various studies use real-time datasets or developed hardware for VODT, none of them have been analyzed in a real-time environment in the practical world. It is also noted that attention-based approaches offer higher accuracy; however, they have higher processing demand, which impedes catering to real-time demands.
- Other critical challenges: The existing studies carried out towards event detection don't deal with data variability as well as scalability. Further, studies considering recognition of human activity and detection of anomalies are highly dependent on orientation and viewpoints which always vary and such challenges are not addressed in object detection. Identification of real-time object detection and tracking are also quite challenging to be accomplished because of same reason.

From the outcome of the above-mentioned research gap associated with existing video object detection methodologies, it is noted that existing schemes do have their own advantages as well as significant limiting attributes. However, from a global context viewpoint, there are certain contributory suggestions which the proposed study has arrived towards improving or addressing the above-mentioned research gaps. Following are some contributory suggestions towards leveraging the performance of existing video object detection methodologies:

- Encouraging consideration of diverse and large-scale training data: Different challenges associated with backgrounds and wide ranges of classes of an object can be effectively addressed considering diverse dataset with better representation. The idea is to achieve better performance on real-time events and effective generalization.
- Need focus on architecture design: Majority of existing studies are use-case specific with not much emphasizing over architecture-based deployment and extending its applicability. The issues of computational

burden and accuracy can be well balanced if architectures are subjected to optimization.

- Needs practical predictive approach: There are various existing predictive models using branches of artificial intelligence (machine learning and deep learning). However, sustainability of such model over extensive test-cases is not verified. Hence, there is a need of a learning model that can adapt to dynamically changing condition is highly demanded. More research towards ensemble approaches can offer extended robustness and minimize system biases.
- Inclusion of contextual information: The problems associated with accuracy of localization of an object as well as uncertainty connected with instances of an object can be well addressed by including contextual information. Extraction of contextual information can be carried out from semantic segmentation, understanding the scene, and also from surrounding objects.

Hence, all the above-mentioned points are contributory suggestion towards improving video object detection methodology. The next section summarizes the essential highlights of this review work.

VII. CONCLUSION

This paper offers an overview of some notable techniques and approaches towards the VODT, which plays one of the crucial intrinsic operations within any surveillance system. Prior to draw conclusive remarks, it is to be noted that proposed review work formulates certain research questions prior to undertake the review work as follows:

- Ro1: Is the existing review paper towards VODT informative enough to draw conclusive remarks about strength and weakness?
- Ro2: What are the frequently used approaches towards VODT?
- Ro3: What are the issues in multiple approaches towards VODT?
- Ro4: How to improve the performance of VODT differently from existing studies.

After reviewing the existing studies both on perspective of implementation and review work, it is noted that existing papers cannot be used for withdrawing conclusive remarks as they are highly symptomatic in nature. It will mean that they deal with narrowed set of issues while leaving other associated issues unaddressed. This is the response towards Ro₁. The response to Ro₂ is that attention-based and tracking-based approaches are most effectively proven and hence frequently adopted. Towards the response for Ro₃, the multiples issues have been presented in the form of research gap discussed in prior section. The similar sub-section of research gap are also essential points that are required to be considered to improve the performance of VODT that are not presented in existing studies. This acts as response for Ro₄. After reviewing various approaches, it is noted that approaches toward VODT are highly specific to the use cases. The studies are mainly not

emphasized in the methodologies. At the same time, more inclination of existing approaches is found towards solving the problems associated with unique use cases, e.g., satellite-based VODT, unmanned aerial vehicle-based VODT, remote sensing-based studies, multiple tracking-based studies, etc. A smaller number of generalized frameworks can address all the issues about high performance and robust tracking and Detection. The analysis also found that the implication of deep learning, Siamese network, and convolution neural networks is constantly rising to improve Detection and tracking performance. Looking into the progress in the study, it can be concluded that there is a long way to go to witness a more high-performance VODT approach. There is a need for more extensive analysis to expose the VODT approach to challenging scenarios of the practical world. In contrast, the existing approaches are too narrow and confined to their adopted research environment. The contribution and novelty of this paper is

1) Unlike existing review work, the current paper captures maximum deployment area, use-cases, and wider variants methodologies adopted towards improving performance of VODT.

2) This paper discusses the new classification of approaches of VODT concerning discreet use cases not reported in existing review work,

3) the paper also introduces a compact discussion associated with the commercial application of VODT, where it can be seen that progress made by commercial application and research work has a wide gap,

4) the paper discusses all the notable research contributions concerning the problems being addressed, the methodology being adopted, their respective strength and weakness,

5) the paper presents a discussion of research trends of VODT to find that evolving approaches and their frequencies of usage,

6) the paper outlines the research gap to signify the importance of enhancing the existing studies for developing high-performance VODT.

The future work will be in the direction of addressing the identified research gap from this review work. For this purpose, the initial work direction will be developing a robust detection module considering the challenging scene context of the video feed. Upon accomplishing a satisfactory detection module assessed over different extensive test environments, the next work will be carried out towards tracking operation. As tracking operation is seamless, both spatial and temporal attributes will be considered for mathematical modeling of the VODT approach that can lead to better optimization of its performance. The sole motive of future work will be to accomplish high-performance VODT in a cost-effective computational approach.

REFERENCES

- [1] M. H Kolekar, Intelligent Video Surveillance Systems-An Algorithmic Approach, CRC Press, ISBN: 9781351649902, 1351649906, 2018
- [2] E. Maiettini, G. Pasquale, L. Rosasco, L. Natale, "Online object detection: a robotics challenge," ACM-Autonomous Robots, Vol.44,

- No.5, pp.739–757, 2020. DOI: <https://doi.org/10.1007/s10514-019-09894-9>
- [3] B. Kaur, S. Singh, "Object Detection using Deep Learning: A Review," ACM- Proceedings of the International Conference on Data Science, Machine Learning and Artificial Intelligence, pp.328–334, 2021. DOI:<https://doi.org/10.1145/3484824.3484889>
- [4] A. Boukerche, Z. Hou, "Object Detection Using Deep Learning Methods in Traffic Scenarios," ACM Computing Surveys, Vol.54, Issue 2, Article No.:30, pp.1–35, 2022. DOI: <https://doi.org/10.1145/3434398>
- [5] Y. Ji, P. Yin, X. Sun, K. H. H. B. Ghazali, N. Guo, "A Comparative Study and Simulation of Object Tracking Algorithms," ACM-The 4th International Conference on Video and Image Processing, pp.161–167, December 2020. DOI: <https://doi.org/10.1145/3447450.3447476>
- [6] Y. Wang, J-N Hwang, G. Wang, H. Liu, K-J Kim, H-M Hsu, J. Cai, H. Zhang, Z. Jiang, R. Gu, "ROD2021 Challenge: A Summary for Radar Object Detection Challenge for Autonomous Driving Applications", ACM-Proceedings of the 2021 International Conference on Multimedia Retrieval, pp.553–559, August 2021. DOI: <https://doi.org/10.1145/3460426.3463658>
- [7] E. Arulprakash, M. Aruldoss, "A study on generic object detection with emphasis on future research directions," ACM-Journal of King Saud University - Computer and Information Sciences, Vol.34, Issue 9, pp 7347–7365, Oct 2022. DOI: <https://doi.org/10.1016/j.jksuci.2021.08.001>
- [8] A. Salari, A. Djavadifar, X. Liu, H. Najjaran, "Object recognition datasets and challenges: A review," ACM-Neurocomputing, Vol.495, Issue C, pp.129–152, Jul 2022. DOI: <https://doi.org/10.1016/j.neucom.2022.01.022>
- [9] İ. Delibaşoğlu, "Surveillance with UAV Videos", Intechopen-Intelligent Video Surveillance - New Perspectives, 2022. DOI: 10.5772/intechopen.105959
- [10] H. Li, Y. Dong, L. Xu, S. Zhang, & J. Wang, "Object detection method based on global feature augmentation and adaptive regression in IoT," SpringerOpen-Neural Computing and Applications, vol.33, pp.4119–4131, 2021
- [11] J-P Mercier, M. Garon, P. Giguère, J-F Lalonde, "Deep Template-based Object Instance Detection", arXiv- Computer Vision and Pattern Recognition, 2019. DOI: <https://doi.org/10.48550/arXiv.1911.11822>
- [12] Zhu, J.; Wang, Z.; Wang, S.; Chen, S. Moving Object Detection Based on Background Compensation and Deep Learning. *Symmetry* 2020, *12*, 1965. <https://doi.org/10.3390/sym12121965>
- [13] A. S. Patel, R. Vyas, O. P. Vyas, M. Ojha, V. Tiwari, "Motion-compensated online object tracking for activity detection and crowd behavior analysis," SpringerOpen-The Visual Computer, 2022
- [14] K-P Kortmann, J. Zumsande, M. Wielitzka, T. Ortmaier, "Temporal Object Tracking in Large-Scale Production Facilities using Bayesian Estimation," Elsevier-IFAC-PapersOnLine, Vol.53, No.2, pp.11125-11131, 2022. DOI: <https://doi.org/10.1016/j.ifacol.2020.12.271>
- [15] A. Mishra, S. Lee, D. Kim, S. Kim, "In-Cabin Monitoring System for Autonomous Vehicles", *Sensors*, vol.22, No.4360, 2022. DOI: <https://doi.org/10.3390/s22124360>
- [16] V. Paidi, H. Fleyeh, J. Håkansson, and R. G. Nyberg, "Tracking Vehicle Cruising in an Open Parking Lot Using Deep Learning and Kalman Filter," Hindawi-Journal of Advanced Transportation, Article ID 1812647, DOI:<https://doi.org/10.1155/2021/1812647>
- [17] T. Nguyen, C. Pham, K. Nguyen, M. Hoai, "Few-shot Object Counting and Detection," arXiv, Computer Vision and Pattern Recognition, 2022. DOI: <https://doi.org/10.48550/arXiv.2207.10988>
- [18] J.S. Murthy, G. M. Siddesh, W-C Lai, B. D. Parameshachari, S.N. Patil, and K. L. Hemalatha, "ObjectDetect: A Real-Time Object Detection Framework for Advanced Driver Assistant Systems Using YOLOv5", Hindawi-Wireless Communication and Mobile Computing, Article ID 9444360, 2022, DOI:<https://doi.org/10.1155/2022/9444360>
- [19] S. Dokania, A. H. A. Hafez, A. Subramanian, M. Chandraker, C.V. Jawahar, "IDD-3D: Indian Driving Dataset for 3D Unstructured Road Scene", arXiv:2210.12878v1 [cs.CV] 23 October 2022
- [20] L. Malburg, M-P Rieder, R. Seiger, P. Klein, R. Bergmann, "Object Detection for Smart Factory Processes by Machine Learning", Elsevier-Procedia Computer Science, Vol.184, pp.581-588, 2021. DOI: <https://doi.org/10.1016/j.procs.2021.04.009>
- [21] A-A Tulbure, A-A Tulbure, E-H Dulf, "A review on modern defect detection models using DCNNs – Deep convolutional neural networks," ScienceDirect-Journal of Advanced Research Vol.35, pp.33-48, 2022. DOI: <https://doi.org/10.1016/j.jare.2021.03.015>
- [22] L. Zhou, L. Zhang, N. Konz, "Computer Vision Techniques in Manufacturing", TechRxiv Preprint, 2021. DOI: <https://doi.org/10.36227/techrxiv.17125652.v2>
- [23] V. Isailovic, A. Peulic, M. Djapan, M. Savkovic, A. M. Vukicevic, "The compliance of head-mounted industrial PPE by using deep learning object detectors," Scientific Reports, vol.12, Article number: 16347, 2022.
- [24] J. Wen, T. Abe, and T. Suganuma, "A Customer Behavior Recognition Method for Flexibly Adapting to Target Changes in Retail Stores," Sensors, vol. 22, no. 18, p. 6740, Sep. 2022, doi: 10.3390/s22186740.
- [25] Y.-S. Yoo, S.-H. Lee, and S.-H. Bae, "Effective Multi-Object Tracking via Global Object Models and Object Constraint Learning," Sensors, vol. 22, no. 20, p. 7943, Oct. 2022, doi: 10.3390/s22207943
- [26] J. M. R. Andaur, G. A. Ruz, and M. Goycoolea, "Predicting Out-of-Stock Using Machine Learning: An Application in a Retail Packaged Foods Manufacturing Company," Electronics, vol. 10, no. 22, p. 2787, Nov. 2021, doi: 10.3390/electronics10222787.
- [27] K. Xia et al., "An Intelligent Self-Service Vending System for Smart Retail," Sensors, vol. 21, no. 10, p. 3560, May 2021, doi: 10.3390/s21103560.
- [28] E. Maltezos et al., "A Video Analytics System for Person Detection Combined with Edge Computing," computation, vol. 10, no. 3, p. 35, Feb. 2022, doi: 10.3390/computation10030035
- [29] Z. Zhang, C. Wang, J. Song, and Y. Xu, "Object Tracking Based on Satellite Videos: A Literature Review," Remote Sensing, vol. 14, no. 15, p. 3674, Jul. 2022, doi: 10.3390/rs14153674.
- [30] S. Chen et al., "Vehicle Tracking on Satellite Video Based on Historical Model," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 7784-7796, 2022, doi: 10.1109/JSTARS.2022.3195522.
- [31] Z. Hu, D. Yang, K. Zhang, and Z. Chen, "Object Tracking in Satellite Videos Based on Convolutional Regression Network With Appearance and Motion Features," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 783-793, 2020, doi: 10.1109/JSTARS.2020.2971657.
- [32] F. Shi, F. Qiu, X. Li, Y. Tang, R. Zhong, and C. Yang, "A Method to Detect and Track Moving Airplanes from a Satellite Video," Remote Sensing, vol. 12, no. 15, p. 2390, Jul. 2020, doi: 10.3390/rs12152390.
- [33] D. Wu, H. Song, and C. Fan, "Object Tracking in Satellite Videos Based on Improved Kernel Correlation Filter Assisted by Road Information," Remote Sensing, vol. 14, no. 17, p. 4215, Aug. 2022, doi: 10.3390/rs14174215.
- [34] S. Xuan, S. Li, M. Han, X. Wan and G. -S. Xia, "Object Tracking in Satellite Videos by Improved Correlation Filters With Motion Estimations," in IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 2, pp. 1074-1086, Feb. 2020, doi: 10.1109/TGRS.2019.2943366.
- [35] Y. Zhang, D. Chen, and Y. Zheng, "Satellite Video Tracking by Multi-Feature Correlation Filters with Motion Estimation," Remote Sensing, vol. 14, no. 11, p. 2691, Jun. 2022, doi: 10.3390/rs14112691.
- [36] Z. Zhou, S. Li, W. Guo, and Y. Gu, "Few-Shot Aircraft Detection in Satellite Videos Based on Feature Scale Selection Pyramid and Proposal Contrastive Learning," Remote Sensing, vol. 14, no. 18, p. 4581, Sep. 2022, doi: 10.3390/rs14184581
- [37] K. Zhu et al., "Single Object Tracking in Satellite Videos: Deep Siamese Network Incorporating an Interframe Difference Centroid Inertia Motion Model," Remote Sensing, vol. 13, no. 7, p. 1298, Mar. 2021, doi: 10.3390/rs13071298.
- [38] Y. You, J. Cao, and W. Zhou, "A Survey of Change Detection Methods Based on Remote Sensing Images for Multi-Source and Multi-Objective Scenarios," Remote Sensing, vol. 12, no. 15, p. 2460, Jul. 2020, doi: 10.3390/rs12152460.
- [39] L. Lei and D. Guo, "Multitarget Detection and Tracking Method in Remote Sensing Satellite Video," Hindawi-Computational Intelligence

- and Neuroscience, Article ID 7381909, 2021. DOI:https://doi.org/10.1155/2021/7381909
- [40] Q. Lin, "Real-Time Multitarget Tracking for Panoramic Video Based on Dual Neural Networks for Multisensor Information Fusion," *Hindawi-Mathematical Problems in Engineering*, Article ID 8313471, 2022. DOI:https://doi.org/10.1155/2022/8313471
- [41] T.J. Ma, "Remote sensing detection enhancement," *Springer-Journal of Big Data*, vol.8, No.127, 2021. DOI: https://doi.org/10.1186/s40537-021-00517-8
- [42] G. Tochon, J. Chanussot, M. Dalla Mura and A. L. Bertozzi, "Object Tracking by Hierarchical Decomposition of Hyperspectral Video Sequences: Application to Chemical Gas Plume Tracking," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4567-4585, Aug. 2017, doi: 10.1109/TGRS.2017.2694159.
- [43] B. UzKent, M. J. Hoffman and A. Vodacek, "Integrating Hyperspectral Likelihoods in a Multidimensional Assignment Algorithm for Aerial Vehicle Tracking," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 9, pp. 4325-4333, Sept. 2016, doi: 10.1109/JSTARS.2016.2560220.
- [44] B. UzKent, A. Rangnekar, and M. J. Hoffman, "Tracking in Aerial Hyperspectral Videos Using Deep Kernelized Correlation Filters," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 449-461, Jan. 2019, doi: 10.1109/TGRS.2018.2856370.
- [45] J. Wei, J. Sun, Z. Wu, J. Yang, and Z. Wei, "Moving Object Tracking via 3-D Total Variation in Remote-Sensing Videos," in *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1-5, 2022, Art no. 3506405, doi: 10.1109/LGRS.2021.3077257.
- [46] E. Çintaş, B. Özyer, and E. Şimşek, "Vision-Based Moving UAV Tracking by Another UAV on Low-Cost Hardware and a New Ground Control Station," in *IEEE Access*, vol. 8, pp. 194601-194611, 2020, doi: 10.1109/ACCESS.2020.3033481.
- [47] C. Deng, S. He, Y. Han, and B. Zhao, "Learning Dynamic Spatial-Temporal Regularization for UAV Object Tracking," in *IEEE Signal Processing Letters*, vol. 28, pp. 1230-1234, 2021, doi: 10.1109/LSP.2021.3086675.
- [48] Z. Ding, S. Liu, M. Li, Z. Lian, and H. Xu, "A Blockchain-Enabled Multiple Object Tracking for Unmanned System With Deep Hash Appearance Feature," in *IEEE Access*, vol. 9, pp. 1116-1123, 2021, doi: 10.1109/ACCESS.2020.3046243.
- [49] X. Liang, J. Zhang, L. Zhuo, Y. Li, and Q. Tian, "Small Object Detection in Unmanned Aerial Vehicle Images Using Feature Fusion and Scaling-Based Single Shot Detector With Spatial Context Analysis," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1758-1770, June 2020, doi: 10.1109/TCSVT.2019.2905881.
- [50] F. Lin, C. Fu, Y. He, F. Guo, and Q. Tang, "Learning Temporary Block-Based Bidirectional Incongruity-Aware Correlation Filters for Efficient UAV Object Tracking," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 6, pp. 2160-2174, June 2021, doi: 10.1109/TCSVT.2020.3023440.
- [51] Y. Bühler, L. Meier and C. Ginzler, "Potential of Operational High Spatial Resolution Near-Infrared Remote Sensing Instruments for Snow Surface Type Mapping," in *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 4, pp. 821-825, April 2015, doi: 10.1109/LGRS.2014.2363237.
- [52] Y. Shan, S. Liu, Y. Zhang, M. Jing, and H. Xu, "LMD-TShip*: Vision Based Large-Scale Maritime Ship Tracking Benchmark for Autonomous Navigation Applications," in *IEEE Access*, vol. 9, pp. 74370-74384, 2021, doi: 10.1109/ACCESS.2021.3079132.
- [53] X. Xue, Y. Li, X. Yin, C. Shang, T. Peng, and Q. Shen, "Semantic-Aware Real-Time Correlation Tracking Framework for UAV Videos," in *IEEE Transactions on Cybernetics*, vol. 52, no. 4, pp. 2418-2429, April 2022, doi: 10.1109/TCYB.2020.3005453.
- [54] J. Ye, C. Fu, F. Lin, F. Ding, S. An, and G. Lu, "Multi-Regularized Correlation Filter for UAV Tracking and Self-Localization," in *IEEE Transactions on Industrial Electronics*, vol. 69, no. 6, pp. 6004-6014, June 2022, doi: 10.1109/TIE.2021.3088366.
- [55] Q. Yu, B. Wang, and Y. Su, "Object Detection-Tracking Algorithm for Unmanned Surface Vehicles Based on a Radar-Photoelectric System," in *IEEE Access*, vol. 9, pp. 57529-57541, 2021, doi: 10.1109/ACCESS.2021.3072897.
- [56] S. Banerjee, H. H. Chopp, J. G. Serra, H. T. Yang, O. Cossairt, and A. K. Katsaggelos, "An Adaptive Video Acquisition Scheme for Object Tracking and Its Performance Optimization," in *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17227-17243, 1 August 1, 2021, doi: 10.1109/JSEN.2021.3081351.
- [57] J. Chen, H. -W. Huang, P. Rupp, A. Sinha, C. Ehmke, and G. Traverso, "Closed-Loop Region of Interest Enabling High Spatial and Temporal Resolutions in Object Detection and Tracking via Wireless Camera," in *IEEE Access*, vol. 9, pp. 87340-87350, 2021, doi: 10.1109/ACCESS.2021.3086499.
- [58] L. Cheng, J. Wang, and Y. Li, "ViTrack: Efficient Tracking on the Edge for Commodity Video Surveillance Systems," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 3, pp. 723-735, 1 March 2022, doi: 10.1109/TPDS.2021.3081254.
- [59] H. Gu et al., "A Collaborative and Sustainable Edge-Cloud Architecture for Object Tracking with Convolutional Siamese Networks," in *IEEE Transactions on Sustainable Computing*, vol. 6, no. 1, pp. 144-154, 1 Jan.-March 2021, doi: 10.1109/TSUSC.2019.2955317.
- [60] S. J. Lee, S. Lee, S. I. Cho, and S. -J. Kang, "Object Detection-Based Video Retargeting With Spatial-Temporal Consistency," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4434-4439, Dec. 2020, doi: 10.1109/TCSVT.2020.2981652.
- [61] C. -J. Liu and T. -N. Lin, "DET: Depth-Enhanced Tracker to Mitigate Severe Occlusion and Homogeneous Appearance Problems for Indoor Multiple-Object Tracking," in *IEEE Access*, vol. 10, pp. 8287-8304, 2022, doi: 10.1109/ACCESS.2022.3144153.
- [62] M. R. Marshall et al., "3-D Object Tracking in Panoramic Video and LiDAR for Radiological Source-Object Attribution and Improved Source Detection," in *IEEE Transactions on Nuclear Science*, vol. 68, no. 2, pp. 189-202, Feb. 2021, doi: 10.1109/TNS.2020.3047646.
- [63] R. Mostafa, H. Baraka, and A. Bayoumi, "LMOT: Efficient Light-Weight Detection and Tracking in Crowds," in *IEEE Access*, vol. 10, pp. 83085-83095, 2022, doi: 10.1109/ACCESS.2022.3197157.
- [64] B. Ramesh et al., "e-TLD: Event-Based Framework for Dynamic Object Tracking," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 10, pp. 3996-4006, Oct. 2021, doi: 10.1109/TCSVT.2020.3044287.
- [65] W. Ren, X. Wang, J. Tian, Y. Tang, and A. B. Chan, "Tracking-by-Counting: Using Network Flows on Crowd Density Maps for Tracking Multiple Targets," in *IEEE Transactions on Image Processing*, vol. 30, pp. 1439-1452, 2021, doi: 10.1109/TIP.2020.3044219.
- [66] S. Sun, N. Akhtar, H. Song, A. Mian, and M. Shah, "Deep Affinity Network for Multiple Object Tracking," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 104-119, 1 January 2021, doi: 10.1109/TPAMI.2019.2929520.
- [67] H. Zhang and X. Ren, "Robust real-time object tracking system for human-following quadruped robot," in *Lecture Notes in Electrical Engineering*, Singapore: Springer Nature Singapore, 2022, pp. 388-397
- [68] J. Cao et al., "Robust object tracking algorithm for autonomous vehicles in complex scenes," *Remote Sens. (Basel)*, vol. 13, no. 16, p. 3234, 2021, doi: 10.3390/rs13163234
- [69] W. Zhao, M. Deng, C. Cheng, and D. Zhang, "Real-time object tracking algorithm based on Siamese network," *Appl. Sci. (Basel)*, vol. 12, no. 14, p. 7338, 2022, doi: 10.3390/app12147338
- [70] C. Du, M. Lan, M. Gao, Z. Dong, H. Yu, and Z. He, "Real-time object tracking via adaptive correlation filters," *Sensors (Basel)*, vol. 20, no. 15, p. 4124, 2020, doi: 10.3390/s20154124.
- [71] P. D. Chakole, V. R. Satpute, and N. Cheggoju, "Crowd behavior anomaly detection using correlation of optical flow magnitude," *J. Phys. Conf. Ser.*, vol. 2273, no. 1, p. 012023, 2022, doi: 10.1088/1742-6596/2273/1/012023.
- [72] R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, "A new approach for abnormal human activities recognition based on ConvLSTM architecture," *Sensors (Basel)*, vol. 22, no. 8, 2022, doi: 10.3390/s22082946.

- [73] Y. Chang *et al.*, "Video anomaly detection with spatio-temporal dissociation," *Pattern Recognit.*, vol. 122, no. 108213, p. 108213, 2022, doi: 10.1016/j.patcog.2021.108213
- [74] S. W. Yahaya, A. Lotfi, and M. Mahmud, "Detecting anomaly and its sources in activities of daily living," *SN Comput. Sci.*, vol. 2, no. 1, 2021, doi: 10.1007/s42979-020-00418-2.
- [75] Y. Yang, F. Angelini, and S. M. Naqvi, "Pose-driven human activity anomaly detection in a CCTV-like environment," *IET Image Process.*, vol. 17, no. 3, pp. 674–686, 2023, doi: 10.1049/ipr2.12664.
- [76] A.-U. Rehman, H. S. Ullah, H. Farooq, M. S. Khan, T. Mahmood, and H. O. A. Khan, "Multi-modal anomaly detection by using audio and visual cues," *IEEE Access*, vol. 9, pp. 30587–30603, 2021, doi: 10.1109/access.2021.3059519
- [77] C. Benegui and R. T. Ionescu, "Improving the authentication with built-in camera protocol using built-in motion sensors: A deep learning solution," *arXiv [cs.CR]*, 2021. doi: 10.3390/1010000.
- [78] H. Shin, K.-I. Na, J. Chang, and T. Uhm, "Multimodal layer surveillance map based on anomaly detection using multi-agents for smart city security," *ETRI J.*, vol. 44, no. 2, pp. 183–193, 2022, doi: 10.4218/etrij.2021-0395.
- [79] I. R. Dave, C. Chen, and M. Shah, "SPAct: Self-supervised privacy preservation for action recognition," *arXiv [cs.CV]*, 2022. Accessed: Jun. 24, 2023. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/papers/Dave_SPAct_Self-Supervised_Privacy_Preservation_for_Action_Recognition_CVPR_2022_paper.pdf
- [80] P. He *et al.*, "Privacy-preserving object detection," *arXiv [cs.CV]*, 2021. Accessed: Jun. 24, 2023. [Online]. Available: <http://arxiv.org/abs/2103.06587>
- [81] J. Zhang, J. Zhou, J. Guo, and X. Sun, "Visual object detection for privacy-preserving federated learning," *IEEE Access*, vol. 11, pp. 33324–33335, 2023, doi: 10.1109/access.2023.3263533.
- [82] T. Bai, S. Fu, and Q. Yang, "Privacy-preserving object detection with secure convolutional neural networks for vehicular edge computing," *Future Internet*, vol. 14, no. 11, p. 316, 2022, doi: 10.3390/fi14110316.
- [83] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, "ImageNet Large Scale Visual Recognition Challenge", *Int. J. Comput. Vis.*, vol.115, pp.211–252, 2015
- [84] D. Damen *et al.*, "The EPIC-KITCHENS dataset: Collection, challenges and baselines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4125–4141, 2021
- [85] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010.
- [86] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge: A retrospective," *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2015.
- [87] P. Dollar, C. Wojek, B. Schiele and P. Perona, "Pedestrian detection: A benchmark," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 304-311, doi: 10.1109/CVPR.2009.5206631.
- [88] G. Awad *et al.*, "TRECVID 2017: Evaluating ad-hoc and instance video search, events detection, video captioning, and hyperlinking," *Nist.gov*. [Online]. Available: <https://www-nlpir.nist.gov/projects/typubs/tv17.papers/tv17overview.pdf>. [Accessed: 28-Apr-2023].
- [89] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio and T. Serre, "HMDB: A large video database for human motion recognition," *2011 International Conference on Computer Vision*, Barcelona, Spain, 2011, pp. 2556-2563, doi: 10.1109/ICCV.2011.6126543.
- [90] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1725-1732, doi: 10.1109/CVPR.2014.223
- [91] L. Leal-Taixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "MOTChallenge 2015: Towards a benchmark for multi-target tracking," *arXiv [cs.CV]*, 2015
- [92] M. Kristan *et al.*, "The Visual Object Tracking VOT2015 Challenge Results," *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, Santiago, Chile, 2015, pp. 564-586, doi: 10.1109/ICCVW.2015.79.
- [93] Y. Wang, P. -M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth and P. Ishwar, "CDnet 2014: An Expanded Change Detection Benchmark Dataset," *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Columbus, OH, USA, 2014, pp. 393-400, doi: 10.1109/CVPRW.2014.126.
- [94] F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung, "A benchmark dataset and evaluation methodology for video object segmentation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [95] H. Mao, X. Yang, and W. J. Dally, "A delay metric for video object detection: What average precision fails to tell," *arXiv [cs.CV]*, 2019.
- [96] X. Zhu, Y. Xiong, J. Dai, L. Yuan, and Y. Wei, "Deep feature flow for video recognition," *arXiv [cs.CV]*, 2016.
- [97] X. Zhu, Y. Wang, J. Dai, L. Yuan, and Y. Wei, "Flow-guided feature aggregation for video object detection," *arXiv [cs.CV]*, 2017.
- [98] X. Zhu, J. Dai, X. Zhu, Y. Wei, and L. Yuan, "Towards high performance video object detection for mobiles," *arXiv [cs.CV]*, 2018
- [99] F. He, N. Gao, J. Jia, X. Zhao, and K. Huang, "QueryProp: Object query propagation for high-performance video object detection," *Proc. Conf. AAAI Artif. Intell.*, vol. 36, no. 1, pp. 834–842, 2022.
- [100] C. Hetang, H. Qin, S. Liu, and J. Yan, "Impression Network for video object detection," *arXiv [cs.CV]*, 2017.
- [101] A. Dosovitskiy *et al.*, "FlowNet: Learning Optical Flow with Convolutional Networks," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 2758-2766, doi: 10.1109/ICCV.2015.316.
- [102] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A critical review of recurrent neural networks for sequence learning," *arXiv [cs.LG]*, 2015.
- [103] C. Zhang and J. Kim, "Modeling Long- and Short-Term Temporal Context for Video Object Detection," *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 2019, pp. 71-75, doi: 10.1109/ICIP.2019.8802920.
- [104] M. Liu and M. Zhu, "Mobile video object detection with temporally-aware feature maps," *arXiv [cs.CV]*, 2017.
- [105] M. Liu, M. Zhu, M. White, Y. Li, and D. Kalenichenko, "Looking fast and slow: Memory-guided mobile video object detection," *arXiv [cs.CV]*, 2019.
- [106] J. Deng, Y. Pan, T. Yao, W. Zhou, H. Li, and T. Mei, "Relation Distillation Networks for video object detection," *arXiv [cs.CV]*, 2019.
- [107] H. Deng *et al.*, "Object Guided External Memory Network for Video Object Detection," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 6677-6686, doi: 10.1109/ICCV.2019.00678.
- [108] Y. Chen, Y. Cao, H. Hu, and L. Wang, "Memory enhanced global-local aggregation for video object detection," *arXiv [cs.CV]*, 2020.
- [109] H. Wu, Y. Chen, N. Wang, and Z. Zhang, "Sequence Level Semantics Aggregation for video object detection," *arXiv [cs.CV]*, 2019.
- [110] W. Yang, B. Liu, W. Li and N. Yu, "Tracking Assisted Faster Video Object Detection," *2019 IEEE International Conference on Multimedia and Expo (ICME)*, Shanghai, China, 2019, pp. 1750-1755, doi: 10.1109/ICME.2019.00301.
- [111] H. Luo, W. Xie, X. Wang, and W. Zeng, "Detect or track: Towards cost-effective video object detection/tracking," *arXiv [cs.CV]*, 2018.
- [112] H.-U. Kim and C.-S. Kim, "CDT: Cooperative detection and tracking for tracing multiple objects in video sequences," in *Computer Vision – ECCV 2016*, Cham: Springer International Publishing, 2016, pp. 851–86
- [113] H. Mao, T. Kong, and W. J. Dally, "CaTDet: Cascaded tracked detector for efficient object detection from video," *arXiv [cs.CV]*,
- [114] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Detect to track and track to detect," *arXiv [cs.CV]*, 2017

- [115]G. Han, X. Zhang, and C. Li, "Semi-supervised DFF: Decoupling detection and feature flow for video object detectors," in *Proceedings of the 26th ACM international conference on Multimedia*, 2018.
- [116]G. Bertasius, L. Torresani, and J. Shi, "Object detection in video with spatiotemporal sampling networks," *arXiv [cs.CV]*, 2018.
- [117]F. Xiao and Y. J. Lee, "Video object detection with an aligned spatial-Temporal Memory," *arXiv [cs.CV]*,
- [118]K. Kang *et al.*, "T-CNN: Tubelets with convolutional neural networks for object detection from videos," *arXiv [cs.CV]*, 2016.
- [119]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv [cs.CV]*, 2015.
- [120]X. Chen, J. Yu, and Z. Wu, "Temporally Identity-Aware SSD with Attentional LSTM," *arXiv [cs.CV]*, 2018
- [121]C. Guo *et al.*, "Progressive Sparse Local Attention for Video object detection," *arXiv [cs.CV]*, 2019.

Pairwise Test Case Generation using (1+1) Evolutionary Algorithm for Software Product Line Testing

Sharafeldin Kabashi Khatir¹, Rabatul Aduni Binti Sulaiman^{2*}, Mohammed Adam Kunna Azrag^{3*},
Jasni Mohamad Zain⁴, Julius Beneoluchi Odili⁵, Samer Ali Al-Shami⁶

Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, Batu Pahat, Malaysia¹
Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, Batu Pahat, Malaysia^{2,3}
Institute for Big Data Analytics & Artificial Intelligence, University Technology Mara, Shah Alam, Malaysia^{2,3}
Institute for Big Data Analytics & Artificial Intelligence, University Technology Mara, Shah Alam, Malaysia⁴
Institute of Digital Humanities, Anchor University, Lagos, Nigeria⁵
Institute of Technology Management and Entrepreneurship, University Technical Malaysia, Melaka, Malaysia⁶

Abstract—Software product line SPLs, or software product lines, are groups of similar software systems that share some commonalities but stand out from one another in terms of the features they offer. Over the past few decades, SPLs have been the focus of a great deal of study and implementation in both the academic and commercial sectors. Using SPLs has been shown to improve product customization and decrease time to market. Additional difficulties arise when testing SPLs because it is impractical to test all possible product permutations. The use of Combinatorial Testing in SPL testing has been the subject of extensive study in recent years. The purpose of this study is to gather and analyze data on combinatorial testing applications in SPL, apply Pairwise Testing using (1+1) evolutionary algorithms to SPL across four case studies, and assess the algorithms' efficacy using predetermined evaluation criteria. According to the findings, the performance of this technique is superior when the case study is larger, that is, when it has a higher number of features, than when the case study is smaller in scale.

Keywords—SPL; SPL testing; combinatorial testing; pairwise testing; evolutionary algorithm; 1+1 EA

I. INTRODUCTION

Software product line (SPL), which is also called software product line development, is a set of software engineering practices for making similar software systems from a single set of software assets and using the same production method for all of them [1]. In other words, SPL is a group of products that is put together based on a set of features. A Feature Model (FM) decides which products are valid. Most of the time, it's not possible to test all of the products that would come from an SPL [1]. So, a small group of these products must be chosen. It used to be best if it got a good order of products.

Effective testing strategies will help any organization that spends a lot of money on software development. This is a demand in SPL because the proportion of testing costs goes up as the costs of developing each product go down. Due to the large number of ways the base software can be changed, testing an entire product line takes a long time and costs a lot of money [2]. These issues had to do with which platform was the

most efficient and what should be tested in separate products based on how hard it was to test the whole product line.

If testing is seen as a long process, the ability and effectiveness of testing can be improved by making the creation of test cases happen automatically. Even though this is a step in the right direction, more research needs to be done on the SPL testing process because it is impossible to test all of the individual systems that make up a large SPL software system. Test case generation in SPL [3], [4] is based on variation point management. The authors in [5] say that SPL testing is hard, but it would be best if all SPL products were set up correctly. In reality, though, this is hard to do. Large product configurations have made SPL testing difficult to handle, as [6]. In fact, some features can be set up in tens of millions of different ways. Since there is pressure to make test suites for the whole product line smaller, it will be hard to test each product in an SPL and stay on budget [7]. Combinatorial testing and other testing methods can cut down on the number of test suites needed for this, but they don't come without their own problems when it comes to scalability.

An Evolutionary Algorithm (EA) is a type of evolutionary computation, which is an optimization algorithm used in the field of artificial intelligence that is based on a population. In EA, processes like reproduction, mutation, recombination, and selection are used to model how natural evolution works. Even though it started in the early 1960s, EA is still a fairly new and changing field, with most research focusing on how it can be used.

The EA, as the name implies, functions similarly to natural evolution. People believe that the processes of recombination, mutation, and selection make individuals more fit because they adapt to their surroundings. An "EA individual" is a single optimization solution, whereas an "EA population" is a collection of "EA individual" optimization solutions.

Combinatorial testing is a type of testing that can be used to test a software product in a thorough way [8]. The goal is to have a product that doesn't have any bugs and can work with a wide range of inputs. Pairwise testing, also called "all-pairs

testing," is a way to check the quality of software by comparing the actual results to what was expected. Here, software testers look at all possible pairs of parameters used to test a feature and compare and contrast them.

Pairwise testing has become an important technique for any software tester over the past few years. This method has been around for almost 20 years, but it has become more popular in the last five. At least 20 tools that can make pairwise test cases have had their information made public up to this point [7, 8].

Based on the examined research, there are three main problems that sum up the issues at hand. First, there are so many possible industrial SPL products that it would be impractical to check each one individually to see if it meets each criterion. Second, it is not possible to do a full test that looks at every possible combination of parameters and values. Lastly, a way for evaluating the effectiveness of EA and creating valid comparisons, as well as a technique for reducing testing effort and shortening testing time using the suggested strategy [9]. This study's objectives are to apply the (1+1) evolutionary algorithm to generate pairwise test cases in software product line testing and to compare the effectiveness of the (1+1) EA in terms of pairwise coverage, execution time, test suite size, and test case redundancy for the mobile phone, vending machine, online shopping, and IoT device case studies.

However, it is imperative to acknowledge that the evaluation of security and privacy-related testing concerns should also be taken into account while conducting SPL testing [50]. Access control and authorization procedures are essential elements of software systems that are responsible for managing sensitive data or executing crucial operations. The process of testing these mechanisms across various products within a software product line presents challenges due to the potential variations in access control requirements and authorization procedures among different products.

In order to address this matter, it is possible to utilize sophisticated testing methodologies, such as pairwise testing, to build a set of test cases that encompass various combinations of access control criteria and authorization processes. The (1+1) evolutionary algorithm can be utilized as an efficient method for generating test cases. The efficiency of these test cases can then be assessed by utilizing metrics such as coverage and defect detection rate.

In addition to access control and authorization methods, software product line testing should also encompass additional important security and privacy-related testing concerns, including but not limited to data privacy, encryption, and secure communication. By effectively addressing these concerns [48, 49], software developers may guarantee security and ensure the privacy of their products.

Segment Particle Swarm Optimization (Se-PSO), Enhanced Segment Particle Swarm Optimization (Ese-PSO), and African Buffalo Optimization (ABO) are swarm intelligence-based optimization algorithms inspired by the behavior of social organisms such as ants, bees, and honeybees. They have been successfully applied to various optimization problems and can also be used for test case generation. These algorithms can be

utilized in pairwise test case generation to generate optimal test cases that cover all possible pairwise combinations of input parameters [10, 11, 12, 13, 14, 47].

Pairwise testing is effective at reducing the number of test cases required for high coverage, but it may not be adequate for testing non-functional requirements such as performance, security, and usability. Furthermore, it has been observed in certain research that the effectiveness of pairwise testing could be reliant upon the particular attributes of the software product line under examination, including the number of features and the level of variability. Hence, additional research is required to assess the efficiency of pairwise testing across various scenarios and to develop more sophisticated methodologies for evaluating software product lines.

The intention of this study is to apply the (1+1) evolutionary algorithm to the task of generating pairwise test cases for software product line testing and to assess the efficacy of this technique using a number of measures, including pairwise coverage, execution time, test suite size, and test case redundancy. The objective is to demonstrate that the (1+1) evolutionary algorithm can be helpful for producing high-quality test cases for software product lines and to encourage the evolution of more sophisticated testing methods for software product lines.

The rest of this article is organized as follows: Section II outlines the works that are related to this study. Section III demonstrates the method for conducting this study with case studies that are used to carry out the experiment. In Section IV, a number of experiments are carried out, and the findings are thoroughly examined. Section V contains a discussion of the study's findings. Finally, in Section VI and VII, we provide a brief summary of the paper and discuss future work respectively.

II. RELATED WORKS

SPL is a collection of software-heavy systems with a common base and features tailored to a specific audience or mission. Features identify SPL members by highlighting shared and unique traits. Feature models express feature relationships and limitations to reflect all SPL outputs.

The SPL testing process is difficult. Testing every product is impractical. The number of configurations (or products) caused by an FM usually grows exponentially with the number of features, resulting in millions of potential products to test. Test engineers are trying to reduce their test suites to meet budgets and deadlines [15].

Software testing a product line takes time [16, 3]. A product line lets a buyer build a software system with many options [17]. Business is embracing SPL. Bosch, Philips, Siemens, General Motors, Hewlett-Packard, Boeing, and Toshiba use product-line approaches to reduce development and maintenance costs, improve quality, and speed product development [17, 18].

A. Software Product Line Testing

Testing software product lines is important because one bug can affect hundreds of thousands or millions of products. The study [19] lists several product line assessment methods.

Product-by-product testing begins by generating and testing concrete products one at a time using single-product testing methods. Family-based testing checks if all products in a family meet the requirement [20].

A family-based approach tests multiple products. Computer simulations represent all line products. By superimposing all product test specifications, modern family-based testing methods don't allow for good testing of software's interaction with hardware and the environment [21, 22]. Family-based testing may be time-consuming and incomplete due to its complex execution environment.

Testers and software engineers use FM to compare and create testable products. Testing all product feature combinations is not always possible. Application complexity reduces product selection. Combinatorial testing is used in the selection process to examine multiple variables. This selection method disregards FM defects. A fault-based approach like mutation-based testing can improve error detection and SPL product compliance. The research [23] suggests mutating products for SPL feature testing. The method can be used to create and evaluate test cases like a test criterion. FM's model features and connections. Feature diagrams (FDs) usually show the FM as a tree.

B. Test Case Generation Approaches in SPL

SPL test case generation has led to several testing methods. Combinatorial and model-based testing are examples. Combinatorial testing prevents tests from growing exponentially by trying all possible input permutations. Combinatorial testing addresses test selection from the whole combinatorial product since testing often has a finite test budget and exhaustive testing is usually intractable. Pairwise combinatorial testing is common here. A family of products with all FM valid pairs of features is the goal [24]. Counting covered pairs which help evaluate the product set.

All-pairs testing, also called pairwise testing, is a way to test software by giving it as many possible combinations of two inputs. This method helps us understand how inputs interact, improving product quality and dependability. Pairwise testing is useful for testing software product lines, which are collections of configurable products. Pairwise testing can improve product line testing and be applied to a case study [25]. Pairwise testing with other methods and business knowledge may reduce testing costs and improve quality [26, 37]. The paper recommends pairwise testing to reduce test cases.

T-wise testing checks all input value permutations with a constraint of "T" inputs. This strategy can help test too many inputs. T-wise testing balances test case volume and coverage [27]. SPL's model-based t-wise testing creates a TS with comprehensive t-wise coverage. A valid t-set has t features that meet some constraints.

Covering arrays in software product line testing improves system failure detection [28]. For testing, a two-layer covering array is used to represent equivalence classes and compute their names in the second layer. Covering arrays are used to test component interactions in a systematic manner. Let N , t , k , and v be integers with $k \geq t \geq 2$ and $v \geq 2$. A covering array

$CA(N; t, k, v)$ is an $N \times k$ array A in which each entry is from a different alphabet, and there is a row of B that equals x for every $N \times t$ subarray B of A and every $x \in \Sigma^t$. Then t denotes the covering array's strength, k the number of factors, and v the number of levels [29].

Model-based testing (MBT) automates test case creation for SPL testing. A Systematic Literature Review (SLR) on MBT for SPL testing is presented by [30]. MBT in SPL issues, evaluation, and solutions are discussed. The study summarizes SPL MBT perspectives in a taxonomic structure. The latest SPL development is taxonomy based MBT classification.

Reduced testing is needed when resources are limited. Risk-based testing [31] is popular for system prioritization. Two other factors determine the probability of system entity damage or loss.

SPL regression testing is difficult because it must test every member of a product family after a change. Regression test selection (RTS) selects a subset of regression test cases to lower regression testing costs [32]. In the product line context, each test case can be executed on multiple products that reuse the test case, making SPL regression testing time-consuming and resource-intensive even with RTS. Eliminating unnecessary test case executions helps.

In [33], a suggested method that finds a group of products where running the test case will cover the same sequence of source code statements and give the same testing results, and then filters out the group from the test case's scope.

C. Search-based Techniques for SPL Testing

SPLs are collections of systems that have the same core functionality but are tailored to meet the needs of specific user groups. All products would have to be tested in theory, but that's not possible in practice. Because of this, "interesting" ones can be chosen to focus on using search-based approaches.

Evolutionary computation is a population-based metaheuristic optimization algorithm used in artificial intelligence research. EA simulates natural evolution using reproduction, mutation, recombination, and selection. EAs mimic natural evolution. Recombination, mutation, and selection are thought to increase fitness by adapting individuals to their environment. EA "population" members are optimization solutions. EAs excel at optimization, scheduling, planning, design, and management [34]. Investments, production, distribution, etc. have these issues.

Initially, a theoretical study of the (1+1) EA is presented and discussed. On a population of one, it only employs the mutation operation and an elitist selection method to generate a new generation. Although the (1+1) EA is the simplest evolutionary algorithm, it shares a fundamental principle with all others [35]. The (1+1) EA locates the maximum of a linear function, as proven by a theorem in [28]. The two members of the population at any given iteration are known as the "parent" and the "offspring," hence the name "1+1." For linear function optimization, the (1+1) Evolutionary Algorithm is predicted to take $O(n \ln n)$ time if the mutation rate is of size $(1/n)$.

Differential evolution (DE), Evolution strategy (ES), and Evolutionary Programming (EP) are all examples of other

Evolutionary Algorithms that can produce multiple offspring and compete [36]. As an illustration, the Evolution Strategy allows for the creation and competition of mutants.

Table I provides a summary of some of the existing search-based techniques for testing in SPL. Moreover, the strengths and weaknesses of the technique are provided as well.

TABLE I. SUMMARY OF STUDIED SEARCH-BASED TESTING TECHNIQUE

Technique	Authors	Strengths	Weaknesses
1+1 Evolutionary Algorithm (1+1 EA)	Slowik & Kwasnicka, Zhou et al 2020. [36]	Simplest EA, requires low requirements and it can reach any point in the search space in a single step.	it's not easy to find a good drift function.
Genetic Algorithm (GA)	Rao & Tripathy 2019. [38]	The ability to make exceptional use of parallel computation, simplicity of use, rapid convergence to the global optimum, few necessary control variables.	Do not scale well with complexity, can be quite slow.
Non-dominated Sorting Genetic Algorithm II (NSGA-II)	Hojjati et al., Muhammad Abid Jamil et al 2018. [39]	Demonstrates elitism and is not dependent on any measure of distributivity.	The computational complexity of solving the problem grows in proportion to the size of the problem.
Strength Pareto Evolutionary Algorithm II (SPEA-II)	Jamil et al 2019. [31]	Utilizes a fine-grained fitness assignment strategy and an improved archive truncation technique.	lack of accuracy in its density estimation

III. METHODOLOGY

The methodology for conducting the research includes four stages. The first thing that will be done is an analysis of the software product line online tools (SPLOT), and then in Step 1, the FM for all of the case studies will be prepared. Following that, the pairwise testing will be carried out using the (1+1) EA in Step 2. Evaluation of the parameters that were employed is the third step. Lastly, an in-depth analysis and comparison of the results is carried out. The research methodology is depicted in Fig. 1.

A. Step 1: Prepare Case Studies using Software Product Lines Online Tools (SPLOT)

SPLOT is a Java2 Web app that uses an HTML template engine to make Ajax-based user interfaces for reasoning and configuration. Because it is web-based, you don't have to update it by hand or download any files, and it's easy to share information (for example, through a feature model repository). Automated reasoning and product configuration are SPLOT's two main offerings right now. To this end, reasoning is centered on the automation of crucial debugging tasks like checking the consistency of feature models and spotting the

presence of dead and common features [6]. Measurements of properties like the number of valid configurations and the degree of variability of feature models are also supported by reasoning. Currently, SPLOT supports interactive configuration for product configuration, wherein users decide at a time, and the configuration system automatically propagates those decisions to enforce their consistency.

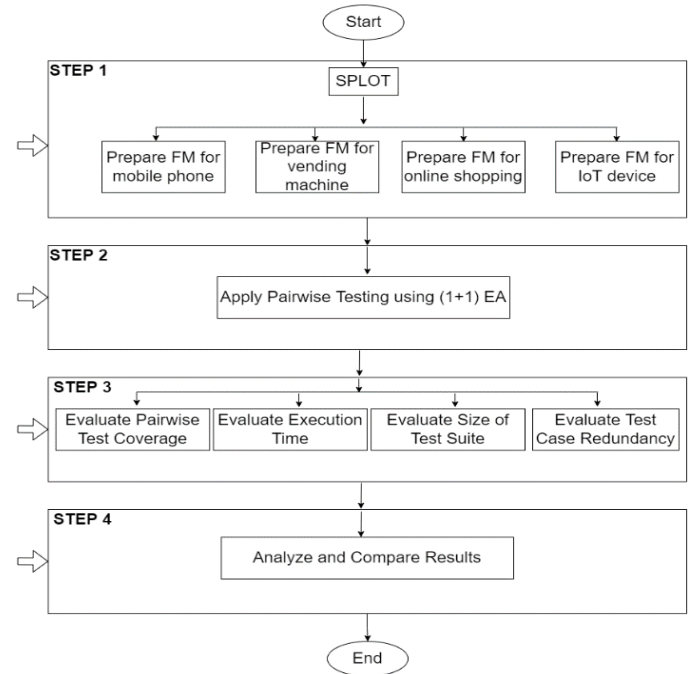


Fig. 1. Research methodology framework.

A major complaint from SPL researchers is the dearth of freely distributed feature models. To address this problem, SPLOT makes available a public model repository with more than 20 genuine models from the literature as well as several automatically generated models with up to 10,000 features each [6].

B. Step 2: Apply Pairwise Testing using (1+1) EA

In this step, we use the PLEDGE tool to generate (1+1)-based pairwise test cases. PLEDGE is a free software program that helps determine which product configurations should be tested in order to cover the most possible combinations of features. Both a command line and a graphical user interface are available for this tool's operation (GUI). All of the following are possible with the current release of PLEDGE:

- FMs loaded from a file: Both the SPLOT and DIMACS (Conjunctive Normal Form) formats are supported by PLEDGE [32].
- Information about FM, such as its limitations and characteristics, can be visualized.
- Making changes to the FM by introducing or removing constraints.
- Producing the test products from the FM by setting parameters for the desired quantity and the time allotted for their production.

- Inputting a list of products and sorting them into a desired order using one of two suggested methods of prioritization.
- Producing a file to store the finalized or prioritized output.

Several settings can be specified to modify the behavior of the 1+1 evolutionary algorithm when using PLEDGE to conduct pairwise testing using the 1+1 evolutionary algorithm.

C. Step 3: Evaluate Testing Parameters

In this stage, all of the parameters, such as pairwise coverage, execution time, the size of test suites, and test case redundancy are evaluated based on the methodology that was utilized in this study.

1) *Pairwise coverage*: For each case study in this research is given a percentage of pairwise coverage for a single run/iteration, which is the discrete combinations of the relevant parameters calculated with PLEDGE. When testing in a black box, pairwise coverage ensures that every possible combination of input parameters is covered by at least one of the test cases. Pairwise testing is more efficient at spotting problems because it is based on the observation that most defects occur due to the interaction of two values. By allowing for systematic testing coverage, pairwise tools can speed up the preparation and implementation stages. By conducting tests in pairs, we can reduce testing time by half without compromising coverage.

The use of k-means and k-medoids clustering techniques in software testing to reduce the test suite and improve the algorithm's performance is discussed by [40, 41]. The technique of pairwise testing is also cited as an efficient means of generating a small test suite with optimal pairwise coverage. Also, [42] proposes a new method for increasing testing efficiency while maintaining testing efficacy. The paper ranks combinatorial test cases according to incremental interaction coverage by repeatedly applying the base choice coverage.

2) *Execution time*: The execution time is the amount of time it takes for a single run to be carried out. Also using PLEDGE, the execution time in seconds is available. For each run, the time taken to finish the run is provided. In the context of pairwise testing, execution time refers to the amount of time required to execute the test cases created using the all-pairs or pairwise testing methodology. The execution time is dependent on variables such as the size of the input parameters, the number of combinations, the performance of the being tested software application, and the testing environment.

Reducing test execution time is crucial for SPL testing, as it enables more efficient testing of product lines and reduces the need for unnecessary testing [43]. Several SPL testing methods have been proposed to decrease test execution redundancy and boost efficiency.

3) *Size of test suite*: A test suite consists of all the test cases that have been logically grouped together. Testing an application to show that it exhibits a certain set of behaviors is

what the test suite is all about. Each test case in a suite will have explicit instructions or goals and details on the system configuration to be used during testing. [41] emphasizes the significance of pairwise testing as a means to circumvent the combinatorial explosion issue. The paper proposes that pairwise testing can be used to test software systems' vast input combinations with fewer test cases. Pairwise testing is presented in [28] as a promising technique with the potential to drastically reduce the number of test cases required for an acceptable level of coverage.

For each case study involved in this study, a different size of test suite can be generated for a single run using PLEDGE. The size of test suites can differ based on the number of features for each case study, a case study with many features is considered big and size of test suites can be high.

4) *Test case redundancy*: This study examines a test suite by finding redundant test cases, which is essential for lowering testing costs. A redundancy score is defined by the redundancy formula, which determines the score by dividing the total number of test cases by the number of duplicates. The redundancy score can be calculated using the formula in Equation 1 below, the total number of redundant test cases is divided by the total number of test cases generated. [44-45] contends that redundancy in test artefacts reduces testing costs.

$$\text{Redundancy Score} = \frac{\sum \text{Redundant test cases}}{\sum \text{Test cases}} \quad (1)$$

D. Step 4: Analyze and Compare the Results

The fourth step is to perform an analysis and comparison of the results, which includes testing and an evaluation of performance. The results of the testing will be analyzed and compared using pairwise coverage, execution time, total test suite size as the criteria. In addition to this, test case redundancy will be calculated, and a graph depicting the findings of the comparison will be offered.

E. Case Studies

Within the scope of this research, four distinct case studies will each be subjected to a pairwise test case generation technique utilizing (1+1) EA. The mobile phone, the vending machine, the online shopping, and the IoT device are the case studies. The reason for choosing the selected case studies is because they are the most used and because they meet the needs to conduct this research, besides, there are many references that has been used those case studies to conduct testing in SPL using other testing techniques. Mobile phone and vending machine case studies are the small case studies in term of number of features. Meanwhile, e-shop and IoT device case studies are the big case studies.

1) *Mobile phone*: The mobile phone industry served as inspiration for the simplified feature model shown in Fig. 2. This model demonstrates how features are incorporated into the process of specifying and developing software for mobile devices, specifically mobile phones. The capabilities of the phone will determine the types of software that can be installed on it. The model stipulates that all mobile devices

must be capable of making and receiving calls as well as displaying data in black-and-white, color, or at a very high resolution on their screens. In addition, the software for mobile phones may, at the user's discretion, include support for satellite navigation systems (GPS) and multimedia devices, such as cameras, MP3 players, or both.

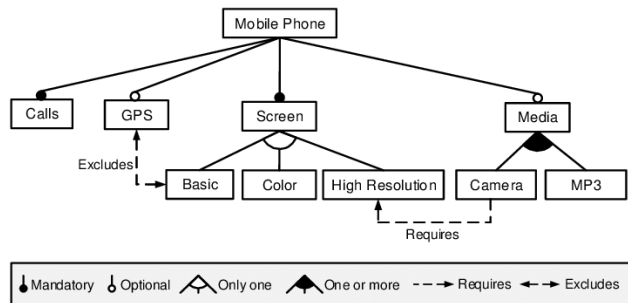


Fig. 2. Feature diagram of product line mobile phone adapted from [38].

2) *Vending machine*: The FD for the snack and drink dispensing machine's SPL is shown in Fig. 3. The vending machine assortment here is formally described by the accompanying feature diagram. Soda, Tea, Free Drinks, and CancelPurchase are used in the feature diagram to represent these products as valid options for the consumer.

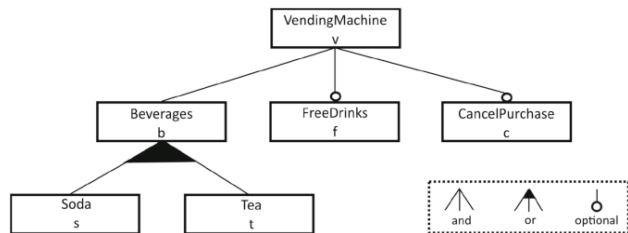


Fig. 3. Feature diagram of product line vending machine adapted from [39].

3) *Online shop*: The feature model that contains information about our online shops is depicted in Fig. 4. The name of the product line is located within the most prominent feature. There are four aspects that are connected to it: The features Catalog, Payment, and Security are connected to the feature that is at the top of the list by arcs that have filled circles at their ends. This indicates that these three features are required, meaning that they are present in each and every product variation. The fact that the Search function is not required is indicated by an arc that terminates in a circle that is not filled in. This descending order of characteristics will continue. For example, the feature Payment includes three sub features: Bank Account, ECoins, and Credit Card. For each product variant, at least one of these sub features must be selected. Both the High and Low sub features of the Security feature are alternative features, which means that only one of them can be selected for each product variant. In addition, there is a textual condition that states that selecting credit cards is only possible when the security level that is being provided is of a high standard.

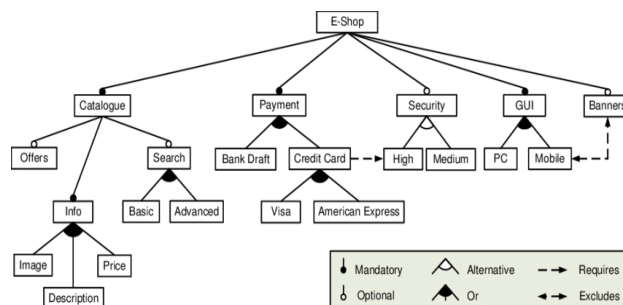


Fig. 4. Feature diagram of product line online shop adapted from [40].

4) *IoT device*: Internet of Things application development is guided by the selection of relevant environmental features and the needs of the end user. An efficient modelling approach, capable of holding all constraints and allowing application development, can be used to control environmental variability. Different uses for the same IoT devices introduce contextual variations that must be managed to ensure efficient development and maximize code reuse. It has been suggested that XML-based feature modelling be used to handle variability management of SPL. Fig. 5 depicts the smart campus IoT system's feature model, complete with predefined relationships and constraints.

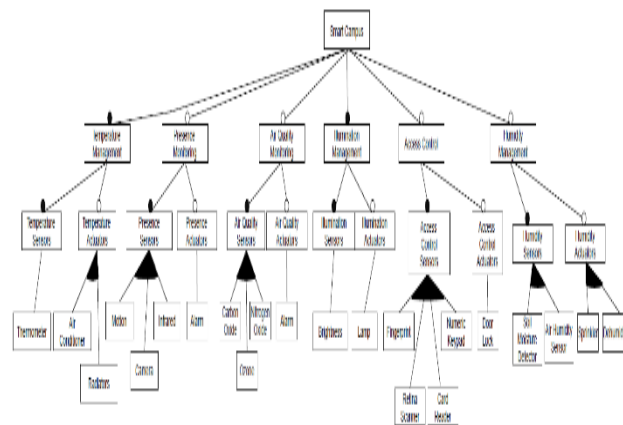


Fig. 5. Smart campus IoT feature diagram adapted from [46].

IV. RESULTS

In this research, four case studies have been tested using the proposed approach. Each case study is tested ten times, and the results of each test case are provided. The four test cases, namely mobile phone, vending machine, online shop, and IoT, are different from each other's, in terms of the number of features, which indicates their size.

Fig. 6 shows the average results of each case study based on the evaluation metrics. For the mobile phone case study, the average test case coverage is 85.23%, and the average execution time is 1.31 seconds. The average number of test cases generated is 2.4, and no test cases are redundant. On the other hand, for the vending machine case study, the averages for the test coverage, execution time, size of the test suite, and percentage of test case redundancy are 75.39%, 1.13 seconds, 2.4, and 23.33%, respectively.

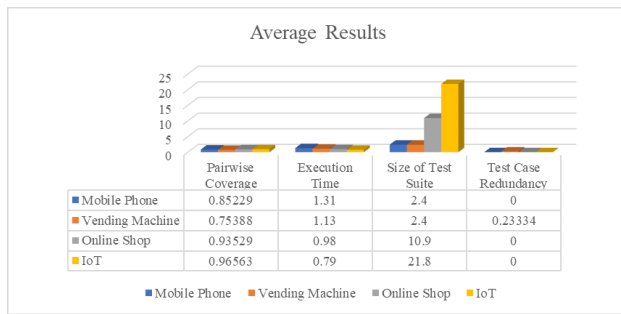


Fig. 6. Average of the results of the case studies.

Moreover, the online shop and smart campus IoT case studies are the big ones in terms of the number of features in this study, and the average percentage of test case coverage is 93.53% for the online shop and 96.56% for the smart campus. Both case study results show the absence of redundant test cases. Meanwhile, the average execution time is 0.98 and 0.79 for the online shop and IoT smart campus, respectively. Lastly, the average size of test cases is 10.9 for the online shop case study and 21.8 for the smart campus case study.

Using PLEDGE to conduct the testing, as shown previously in the mobile phone and vending machine case studies, produced less efficient results compared to the other two case studies. The size of the test suite is smaller because there are fewer features; meanwhile, there are more redundant test cases because, using PLEDGE, the results showed that the tool cannot produce a large number of test suites for small case studies, and the results also showed that the testing takes more time to run, which is remarkably unexpected.

V. DISCUSSION

This study's goals are to apply the (1+1) evolutionary algorithm to generate pairwise test cases in software product line testing with the help of the proposed tool. The effectiveness of the algorithm using four metrics, including pairwise coverage, execution time, size of the test suite, and test case redundancy, for all of the case studies that were chosen, and to conclude that the (1+1) evolutionary algorithm is useful for generating pairwise test cases in software product line testing.

The literature research clarifies the difficulties associated with testing software product lines, mostly stemming from the wide range of possible configurations and the need for effective testing methodologies. The findings of this study indicate that the (1+1) evolutionary algorithm is a successful approach for generating pairs test cases in the context of software product line testing. This algorithm proves to be effective in reducing the number of required test cases while simultaneously obtaining a high level of coverage.

Furthermore, the literature review examines the significance of assessing the efficiency of testing methodologies by considering several factors, including pairwise coverage, execution time, and test suite size. The study presents a comparative analysis of the effectiveness of the (1+1) evolutionary algorithm across four distinct case studies, shedding light on its performance in diverse circumstances.

Moreover, the literature review underscores the necessity for more research and advancement in the domain of software product line testing, encompassing the utilization of search-based algorithms like evolutionary algorithms. The findings and methodology presented in the study make a valuable contribution to the field of research by illustrating the successful performance of the (1+1) evolutionary algorithm in producing pairs test cases for software product line testing.

VI. CONCLUSION

The purpose of this research is to apply the (1+1) evolutionary algorithm using PLEDGE in order to generate pairs of test cases for software product line testing. Using four criteria—pairwise coverage, execution time, test suite size, and test case redundancy—for each of the selected case studies, it was discovered that the 1+1 evolutionary algorithm is beneficial for creating pairwise test cases in software product line testing. When using PLEDGE to conduct the testing, the results demonstrated that this method yields superior results when the case study is large, which means it has a large number of features, compared to when the case study is small.

Among the four case studies, the online shop and IoT case studies achieved good results in comparison to the mobile phone and vending machine case studies. This is because the online shop and IoT case studies have a large number of features; therefore, by using the PLEDGE tool, a good result has been achieved in comparison to when a case study has a small number of features, such as the mobile phone and vending machine case studies.

Online shop and IoT case studies achieved better results than mobile phone and vending machine case studies. The average for pairwise coverage recorded the best for online shop and IoT at 93.53% and 96.56%, respectively. Meanwhile, the average execution time and the size of test suites are noted to be better for online shop and IoT case studies at 0.98s and 0.79s, 10.9s and 21.8s respectively. Also, both case studies showed the absence of redundant test cases. On the other hand, the finding revealed that the mobile phone and vending machine case studies achieved less performance due to the fact that both were considered small with a small number of features.

VII. FUTURE WORKS

From this research, we were able to gather the following list of significant insights regarding areas for possible development and improvement: first, there are numerous ways to produce test cases using the existing software testing tools. Second, the program used in this research, PLEDGE, has problems and must be executed multiple times before producing meaningful results. The production and prioritization of test cases is a crucial aspect of SPL testing, and there has been an increasing trend in recent years to leverage search-based algorithms as a solution strategy.

ACKNOWLEDGMENT

This research was supported by University Tun Hussein Onn Malaysia (UTHM) through Tier 1 Grant (Vot H937).

REFERENCES

- [1] Hierons, R. M., Li, M., Liu, X., Parejo, J. A., Segura, S., & Yao, X. (2020). Many-objective test suite generation for software product lines. *ACM Transactions on Software Engineering and Methodology*, 29(1). <https://doi.org/10.1145/3361146>.
- [2] Cico, O., Jaccheri, L., Nguyen-Duc, A., & Zhang, H. (2021). Exploring the intersection between software industry and Software Engineering education-A systematic mapping of Software Engineering Trends. *Journal of Systems and Software*, 172, 110736.
- [3] Souza, M. R. D. A., Veado, L., Moreira, R. T., Figueiredo, E., & Costa, H. (2018). A systematic mapping study on game-related methods for software engineering education. *Information and software technology*, 95, 201-218.
- [4] Lee, J., Kang, S., & Jung, P. (2020). Test coverage criteria for software product line testing: Systematic literature review. *Information and Software Technology*, 122, 106272.
- [5] Santos, I., Melo, S. M., de Souza, P. S. L., & Souza, S. R. (2019, September). Testing techniques selection: A systematic mapping study. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering* (pp. 347-356).
- [6] Horcas, J. M., Pinto, M., & Fuentes, L. (2019, September). Software product line engineering: a practical experience. In *Proceedings of the 23rd International Systems and Software Product Line Conference-Volume A* (pp. 164-176).
- [7] Wang, R., Artho, C., Kristensen, L. M., & Stolz, V. (2020, December). Multi-objective Search for Model-based Testing. In *2020 IEEE 20th International Conference on Software Quality, Reliability and Security (QRS)* (pp. 130-141). IEEE.
- [8] Al-Hajjaji, M., Thüm, T., Lochau, M., Meinicke, J., & Saake, G. (2019a). Effective product-line testing using similarity-based product prioritization. *Software & Systems Modeling*, 18, 499-521.
- [9] Dominka, S., Mandl, M., Dubner, M., & Ertl, D. (2018). Using combinatorial testing for distributed automotive features: Applying combinatorial testing for automated feature-interaction-testing. *2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, 2018-January*, 490-495. <https://doi.org/10.1109/CCWC.2018.8301632>.
- [10] Kunna, Mohammed Adam, Tuty Asmawaty Abdul Kadir, Muhammad Akmal Remli, Noorlin Mohd Ali, Kohbalan Moorthy, and Noryanti Muhammad. "An enhanced segment particle swarm optimization algorithm for kinetic parameters estimation of the main metabolic model of *Escherichia coli*." *Processes* 8, no. 8 (2020): 963.
- [11] Azrag, Mohammed Adam Kunna, Tuty Asmawaty Abdul Kadir, and Aqeel S. Jaber. "Segment particle swarm optimization adoption for large-scale kinetic parameter identification of *Escherichia Coli* metabolic network model." *IEEE Access* 6 (2018): 78622-78639.
- [12] Azrag, Mohammed Adam Kunna, Jasni Mohamad Zain, Tuty Asmawaty Abdul Kadir, Marina Yusoff, Aqeel Sakhy Jaber, Hybat Salih Mohamed Abdhrman, Yasmeen Hafiz Zaki Ahmed, and Mohamed Saad Bala Husain. "Estimation of Small-Scale Kinetic Parameters of *Escherichia coli* (E. coli) Model by Enhanced Segment Particle Swarm Optimization Algorithm ESe-PSO." *Processes* 11, no. 1 (2023): 126.
- [13] Azrag, Mohammed Adam Kunna, Tuty Asmawaty Abdul Kadir, and Noorlin Mohd Ali. "A Comparison of Particle Swarm optimization and Global African Buffalo Optimization." In *IOP Conference Series: Materials Science and Engineering*, vol. 769, no. 1, p. 012034. IOP Publishing, 2020.
- [14] Odili, Julius Beneoluchi, Mohd Nizam Mohmad Kahar, Shahid Anwar, and Mohammed Adam Kunna Azrag. "A comparative study of African buffalo optimization and randomized insertion algorithm for asymmetric travelling salesman's problem." In *2015 4th International Conference on Software Engineering and Computer Systems (ICSECS)*, pp. 90-95. IEEE, 2015.
- [15] Onipede, S. F., Bashir, N. A., & Abubakar, J. (2022). Small open economies and external shocks: an application of Bayesian global vector autoregression model. *Quality & Quantity*. <https://doi.org/10.1007/s11135-022-01423-8>.
- [16] Zhang, Y., Kong, W., Li, D., & Liu, X. (2020, October). Design and Implementation of Automatic Matching and Remote Screening System for Intelligent Security Inspection. In *Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education* (pp. 76-83).
- [17] Edded, S., Sassi, S. B., Mazo, R., Salinesi, C., & Ghezala, H. B. (2019). Collaborative configuration approaches in software product lines engineering: A systematic mapping study. *Journal of Systems and Software*, 158, 110422.
- [18] Ruland, S., Lochau, M., & Jakobs, M. C. (2020). HybridTiger: Hybrid model checking and domination-based partitioning for efficient multi-goal test-suite generation (competition contribution). *Fundamental Approaches to Software Engineering*, 12076, 520.
- [19] Kolesnikov, S., Siegmund, N., Kästner, C., Grebhahn, A., & Apel, S. (2019). Tradeoffs in modeling performance of highly configurable software systems. *Software & Systems Modeling*, 18, 2265-2283.
- [20] Mesa, O., Vieira, R., Viana, M., Durelli, V. H., Cirilo, E., Kalinowski, M., & Lucena, C. (2018, September). Understanding vulnerabilities in plugin-based web systems: an exploratory study of wordpress. In *Proceedings of the 22nd International Systems and Software Product Line Conference-Volume 1* (pp. 149-159).
- [21] Lee, J., Kang, S., & Jung, P. (2020). Test coverage criteria for software product line testing: Systematic literature review. *Information and Software Technology*, 122, 106272.
- [22] Sulaiman, R. A., Jawawi, D. N., & Halim, S. A. (2023). Cost-effective test case generation with the hyper-heuristic for software product line testing. *Advances in Engineering Software*, 175, 103335.
- [23] Al-Hajjaji, M., Thüm, T., Lochau, M., Meinicke, J., & Saake, G. (2019). Effective product-line testing using similarity-based product prioritization. *Software and Systems Modeling*, 18(1), 499-521. <https://doi.org/10.1007/s10270-016-0569-2>
- [24] Akimoto, H., Isogami, Y., Kitamura, T., Noda, N., & Kishi, T. (2019, December). A prioritization method for spl pairwise testing based on user profiles. In *2019 26th Asia-Pacific Software Engineering Conference (APSEC)* (pp. 118-125). IEEE.
- [25] Wang, Y., Sun, Y., Wu, X., Shanghai cai jing da xue, Tong ji da xue (China), Suzhou da xue, Institute of Electrical and Electronics Engineers. Beijing Section, & Institute of Electrical and Electronics Engineers. (2018). *Proceedings of the 2018 IEEE International Conference on Progress in Informatics and Computing: December 14-16, 2018, Suzhou, China*.
- [26] Morgan, J. (2018). Combinatorial testing: an approach to systems and software testing based on covering arrays. *Analytic methods in systems and software testing*, 131-158.
- [27] Xiang, Y., Huang, H., Member, S., Li, M., Li, S., & Yang, X. (2020). Looking For Novelty in Search-based Software Product Line Testing. In *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*.
- [28] Rabatul Aduni Sulaiman, Dayang Norhayati Abang Jawawi, & Shahliza Abdul Halim. (2022). Classification Trends Taxonomy of Model-based Testing for Software Product Line: A Systematic Literature Review. *KSII Transactions on Internet and Information Systems*, 16(5). <https://doi.org/10.3837/tiis.2022.05.008>
- [29] Jahan, H., Feng, Z., & Mahmud, S. H. (2020). Risk-based test case prioritization by correlating system methods and their associated risks. *Arabian Journal for Science and Engineering*, 45, 6125-6138.
- [30] Jung, P., Kang, S., & Lee, J. (2020). Efficient regression testing of software product lines by reducing redundant test executions. *Applied Sciences (Switzerland)*, 10(23), 1-21. <https://doi.org/10.3390/app10238686>
- [31] Slowik, A., & Kwasnicka, H. (2020a). Evolutionary algorithms and their applications to engineering problems. In *Neural Computing and Applications* (Vol. 32, Issue 16, pp. 12363-12379). Springer. <https://doi.org/10.1007/s00521-020-04832-8>
- [32] Huang, Z., Zhou, Y., Xia, X., & Lai, X. (2020). An improved (1+1) evolutionary algorithm for k-median clustering problem with performance guarantee. *Physica A: Statistical Mechanics and Its Applications*, 539. <https://doi.org/10.1016/j.physa.2019.122992>

- [33] Slowik, A., & Kwasnicka, H. (2020). Evolutionary algorithms and their applications to engineering problems. *Neural Computing and Applications*, 32, 12363-12379.
- [34] Rao, D. S., & Tripathy, D. P. (2019). A genetic algorithm approach for optimization of machinery noise calculations. *Noise and Vibration Worldwide*, 50(4), 112-123. <https://doi.org/10.1177/0957456519839409>.
- [35] Hojjati, A., Monadi, M., Faridhosseini, A., & Mohammadi, M. (2018). Application and comparison of NSGA-II and MOPSO in multi-objective optimization of water resources systems. *Journal of Hydrology and Hydromechanics*, 66(3), 323-329. <https://doi.org/10.2478/johh-2018-0006>.
- [36] Jamil, M. A., Nour, M. K., Alhindi, A., Awang Abhubakar, N. S., Arif, M., & Aljabri, T. F. (2019). Towards Software Product Lines Optimization Using Evolutionary Algorithms. *Procedia Computer Science*, 163, 527-537. <https://doi.org/10.1016/j.procs.2019.12.135>.
- [37] Hierons, R. M., Li, M., Liu, X., Parejo, J. A., Segura, S., & Yao, X. (2020). Many-objective test suite generation for software product lines. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 29(1), 1-46.
- [38] Huang, S., Sun, J., & Feng, Y. (2018). Pairwise covariates-adjusted block model for community detection. *arXiv preprint arXiv:1807.03469*.
- [39] Al-Hajjaji, M., Thüm, T., Lochau, M., Meinicke, J., & Saake, G. (2019b). Effective product-line testing using similarity-based product prioritization. *Software & Systems Modeling*, 18, 499-521.
- [40] Di Silvestro, F. (2020). Improving testing reusability and automation for software product lines.
- [41] Din, F., & Zamli, K. Z. (2019). Pairwise Test Suite Generation Using Adaptive Teaching Learning-Based Optimization Algorithm with Remedial Operator (pp. 187-195). https://doi.org/10.1007/978-3-319-99007-1_18
- [42] Jung, P., Kang, S., & Lee, J. (2020). Efficient regression testing of software product lines by reducing redundant test executions. *Applied Sciences*, 10(23), 8686.
- [43] Ngoumou, A., & Ndjodo, M. F. (2018). Feature-Relationship Models: A Paradigm for Cross-hierarchy Business Constraints in SPL. *International Journal of Computer Science and Information Security (IJCSIS)*, 16(9).
- [44] Dubsloff, C. (2019). Compositional feature-oriented systems. In *Software Engineering and Formal Methods: 17th International Conference, SEFM 2019, Oslo, Norway, September 18-20, 2019, Proceedings 17* (pp. 162-180). Springer International Publishing.
- [45] Sulaiman, R. A. B. (2020). Cost-Effective Model-Based Test Case Generation and Prioritization For Software Product Line (Doctoral dissertation, Universiti Teknologi Malaysia).
- [46] Corradini, F., Fedeli, A., Fornari, F., Polini, A., & Re, B. (2021). FloWare: An Approach for IoT Support and Application Development. *Lecture Notes in Business Information Processing*, 421, 350-365. https://doi.org/10.1007/978-3-030-79186-5_23.
- [47] Wang, Z. J., Yang, Q., Zhang, Y. H., Chen, S. H., & Wang, Y. G. (2023). Superiority combination learning distributed particle swarm optimization for large-scale optimization. *Applied Soft Computing*, 136, 110101. <https://doi.org/10.1016/J.ASOC.2023.110101>.
- [48] Tauqeer, O. B., Jan, S., Khadidos, A. O., Khadidos, A. O., Khan, F. Q., & Khattak, S. (2021). Analysis of security testing techniques. *Intelligent Automation & Soft Computing*, 29(1), 291-306.
- [49] Campanile, L., Iacono, M., & Mastroianni, M. (2022, September). Towards privacy-aware software design in small and medium enterprises. In *2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)* (pp. 1-8). IEEE.
- [50] Stanciu, A. M. (2023). Theoretical Study of Security for a Software Product. In *Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 1* (pp. 233-242). Singapore: Springer Nature Singapore.

Campus Network Intrusion Detection Based on Gated Recurrent Neural Network and Domain Generation Algorithm

Qi Rong^{1*}, Guang Zhao²

Party Committee Organization Department, Jilin Institute of Architecture and Technology, Changchun, China¹
School of Statistics, Shandong University of Finance and Economics, Jinan, China²

Abstract—Network attacks are diversified, rare and Universal generalization. This has made the exploration and construction of network information flow packet threat detection systems, which becomes a hot research topic in preventing network attacks. So this study establishes a network data threat detection model based on traditional network threat detection systems and deep learning neural networks. And convolutional neural network and data enhancement technology are used to optimize the model and improve rare data recognizing accuracy. The experiment confirms that this detection model has a recognition probability of approximately 11% and 42% for two rare attacks when N=1, respectively. When N=2, their probabilities are 52% and 78%, respectively. When N=3, their recognition probabilities are approximately 85% and 92%, respectively. When N=4, their recognition probabilities are about 58% and 68%, respectively, with N=3 having the best recognition effect. In addition, the recognition efficiency of this model for malicious domain name attacks and normal data remains around 90%, which has significant advantages compared to traditional detection systems. The proposed network data flow threat detection model that integrates Gated Recurrent Neural Network and Domain Generation Algorithm has certain practicality and feasibility.

Keywords—Gated recurrent; domain generation algorithm; campus network; threat detection; neural network

I. INTRODUCTION

Technological progress has led to an increasing demand for information technology and a high informatization degree. Therefore, the internet has become an important way for people to improve their efficiency, quality of life, and increase personal income. Internet popularization has made people's lives more and more transparent. People rely heavily on the computer internet in many fields such as daily life, economic management, and financial investment [1]. The global Internet economy has even accounted for 10% of the global GDP, reaching the level of more than ten trillion US dollars. However, while the network has brought us convenience, many malicious attackers will use various means such as Botnet, system vulnerabilities, malicious domain names and trojans to attack, steal, destroy or modify data on the local network without authorization [2]. As of 2022, the losses caused by global cybercrime to people have exceeded \$6 trillion, and these cyber threats are increasing exponentially over time [3]. Currently, campus education networks and research departments are the primary targets of attackers, with each institution or campus network facing an average of thousands

of malicious network attacks per week, an increase of at least 50% compared to 2021 [4]. The existing attack methods for network data have the characteristics of diversification, uniqueness, and fast update. Traditional intrusion detection systems are often unable to effectively identify new and unknown attacks, because they rely on known rules and features [5]. For example, for new attacks such as zero-day vulnerabilities and advanced persistent threats, traditional intrusion detection may not be able to detect and alert in time. And traditional intrusion detection may fail to detect attacks based on covert communication, such as steganography or encrypted communication, and even generate a large number of false positives, wasting time and resources. Therefore, traditional intrusion detection systems have great limitations when facing new attacks, high false positive rate, complex environment and lack of context information. In this context, research attempts to integrate deep learning neural networks and data augmentation techniques into the traditional Network Intrusion Detection System (NIDS) to make it more intelligent and self-learning, to enhance the recognition probability of new rare attack methods.

This study conducted technical exploration and analysis from four aspects. Firstly, the relevant research on the current network DS threat detection system was discussed and summarized. Secondly, the comprehensive applications of Gated Recurrent Neural Network (GRNN), Convolutional Neural Network (CNN), and Domain Generation Algorithm (DGA) were analyzed, including the construction of a network data threat detection system. Then, experimental verification and data comparison analysis were conducted on the data stream (DS) threat detection model of network. Finally, there is a comprehensive overview of entire article and a reflection and summary of its shortcomings.

II. RELATED WORKS

While internet popularization has made people's lives more convenient, network data threats and intrusions have also become increasingly common. Building a DS detection model has become a hot research and exploration field for some experts at home and abroad. Its characteristics include fast and accurate learning, recognition, and fine segmentation of campus network and other DS quantities. Zhou et al. proposed a hierarchical adversarial attack generation method based on the graph neural network (NN) for the intelligent intrusion detection (ID) problem of Internet of Things (IoT), which

improved the system's recognition accuracy against network attacks [6]. The vulnerabilities in IoT are prone to attacks and other issues. In this regard, Nimbalkar and Kshirsagar proposed an attack data detection and comparison system based on the ID system of Feature selection and information gain method, which improved the identification efficiency of denial of service and other attacks [7]. After the automobile network is connected, it is vulnerable to network attacks and accidents. In response, Moubayed et al. proposed a multi-layer hybrid ID system based on feature based ID systems and anomaly based ID systems, thereby improving the recognition efficiency against known and unknown attacks [8]. Guo et al. proposed a spam detection method based on pre trained bidirectional encoder representation and machine learning (ML) algorithm to address the issue of spam detection. Two datasets were used for performance testing in the experiment, confirming that this method effectively improves the detection efficiency of spam [9]. Hidayat et al. proposed a new network ID technology for data detection in network attacks, based on the ML model and integrating multiple technologies, which improved the detection efficiency of network attacks [10].

In addition, Binbusayyis and Vaiyapuri constructed a joint optimization ID framework based on classifiers such as convolutional encoders to address network security issues. This effectively improves the detection ability for unknown attacks [11]. Wang et al. proposed a new ID model for network intrusion based on ML and Decision boundary, integrating popular evaluation methods and ID system models. This increases the recognition probability against boundary attacks [12]. Krishnaveni and others proposed a new attack classification method based on the univariate integration Feature selection technology and classification technology to solve the problem that cloud computing servers are vulnerable to attacks. This improves the detection efficiency of the ID model for attack data [13]. Azizan et al. proposed a new ID method for identifying attack data in large amounts of data, based on ML and incorporating algorithms such as decision jungle. This improves the identification efficiency of attack data in big data [14]. Aimed at the identification problem under multiple attacks on the network, Almomani combines Particle swarm optimization algorithm based on ID system and proposes a new ID model. This effectively improves the recognition efficiency for multiple attacks [15]. Rashid et al. proposed a new ID model based on ML, which integrates tree based stack integration technology to address the classification problem of anomalies and normals. This interference improved the recognition efficiency of network abnormal traffic [16].

From the research from various countries, ID systems have low efficiency in identifying multiple types of network attacks. Most studies only focus on improving ID system's efficiency in identifying attack types. They overlooked the intelligence of ID system and improved learning efficiency for unknown attacks and recognition efficiency for rare attacks. Therefore, the deep detection network model developed using GRNN and DGA has a certain degree of innovation.

III. DESIGN AND IMPLEMENTATION OF NIDS

Unlike traditional threat detection systems, the detection model using GRNN model and DGA fusion has a certain innovation. Therefore, to ensure data detection model's detection accuracy can continuously be deeply refined, the model design and implementation are particularly important. Therefore, this section mainly analyzes the model implementation principle and system construction.

A. Network Threat Detection Model and NN Technology

To conduct real-time monitoring and analysis of DS generated locally on campus network and DS passing through campus local network, NIDS is needed to timely identify network attacks and adopt corresponding strategies [17]. Fig. 1 shows the detection process of this system against DS threats.

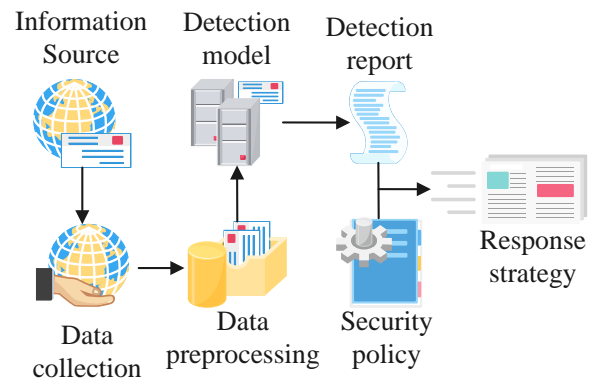


Fig. 1. Flowchart of data flow threat detection.

According to Fig. 1, the entire NIDS detection process mainly includes three steps: information flow data capture, network packet structure content analysis, and abnormal activity strategy response. Among them, data capture mainly utilizes packet capture tools for data capture. Data analysis mainly uses models to make detailed judgments and intelligent decisions. Policy response is mainly aimed at the network abnormal activities by saving the characteristic code, sending information or screen display abnormalities to remind and cooperate with the Network engineer to carry out human intervention to eliminate hidden dangers. Fig. 2 shows the detection logic model of NIDS for network DS.

Through Fig. 2, NIDS accumulates the original feature library based on the initial data and supplements and corrects the feature library through continuous learning and training. Then it makes judgments and decisions on new real-time data on the foundation of vast feature library. Among them, learning historical training experience data is particularly important for NIDS, so deep learning Recurrent Neural Network (RNN) is needed to make NIDS more intelligent [18]. RNN model mainly collects and analyzes the texture characteristics of data through a multi-layer NN composed of simulated neurons. So it can excavate the original feature expressions behind diverse data and make intelligent judgments on richer new data. Fig. 3 is the schematic diagram of RNN.

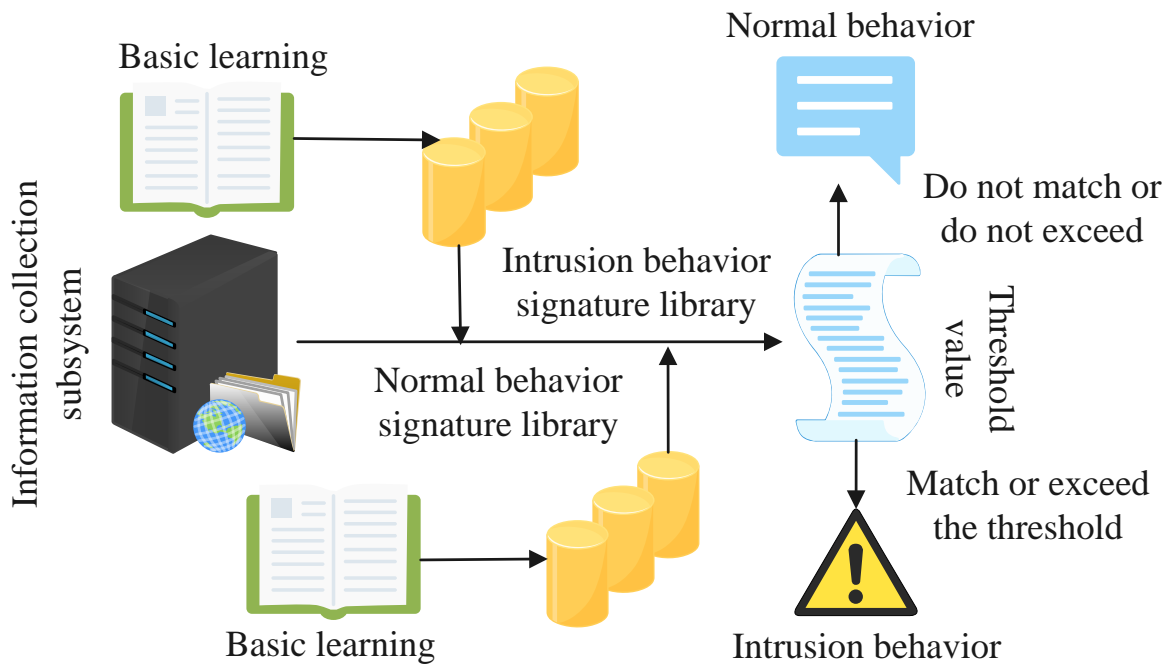


Fig. 2. Data flow detection model.

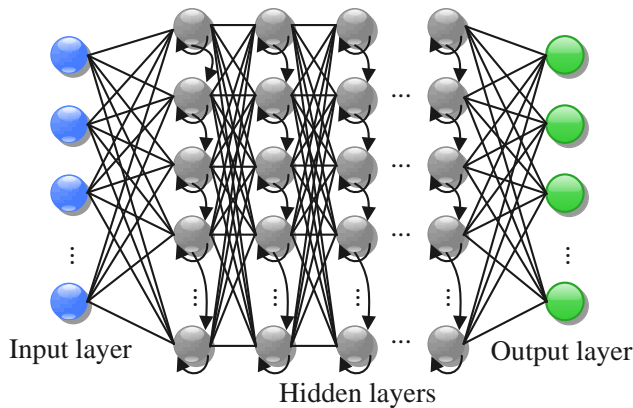


Fig. 3. Recurrent neural network models for deep learning.

From Fig. 3, RNN is a memory feedback bidirectional Transitive model composed of one input layer, multiple hidden layers and one output layer. Each layer of neurons can receive signals from the same or previous layer of neurons and perform nonlinear transformations to accumulate and output to the next layer until the output layer is reached. Due to the tendency of RNN to form gradient decay or gradient explosion when the time difference is large or small, it is difficult for RNN to obtain long-term or short-term dependencies in time series in practical applications. To this end, Gated Recurrent Unit (GRU) should be introduced into RNN to improve its memory unit and better capture long-term or short-term dependencies in time series. Eq. (1) represents its connection relationship [19].

$$\begin{cases} f_t = S(D_{xf}x_t + D_{yf}y_{t-1} + D_{zf}z_{t-1}) \\ i_t = S(D_{xi}x_t + D_{yi}y_{t-1} + D_{zi}z_{t-1}) \\ z_t = f_t H z_{t-1} + i_t H \tanh(D_{xz}x_t + D_{yz}y_{t-1}) \\ O_t = S(D_{xO}x_t + D_{yO}y_{t-1} + D_{zO}z_{t-1}) \\ y_t = O_t H \tanh(z_t) \end{cases} \quad (1)$$

In Eq. (1), f refers to the forgetting gate. S refers to a growth curve function (sigmoid function). D refers to a pending parameter. x refers to the input value. y refers to the output value. z refers to cell state value. t and $t-1$ represent the current and the previous time, respectively. i refers to the input gate. H refers to the image matrix transformation Hadamard matrix. O refers to the output gate. $\tanh()$ refers to a hyperbolic tangent function. In addition, the lack of feedback from neurons themselves is a one-way transmission lacking error adjustment mechanisms. Therefore, to make the gradient descent optimization algorithm and the error Backpropagation work and improve the accuracy of DNN, it is necessary to quantify error information between the estimated and the actual values with cost function. And because NN input values are generally nonlinear discrete values, it is necessary to introduce a linear regression Softmax function to better predict the discrete results. Only by accurately classifying the results according to different weights can more accurate cost function values be obtained. Eq. (2) is the mathematical expression of Softmax value.

$$\left\{ \begin{array}{l} \alpha(x)_j = e^{x_j} / \sum_{k=1}^k e^{x_k}, j=1, \dots, k \\ f_{\beta}(x^j) = \frac{\begin{bmatrix} e^{\beta_0^j x^j} \\ e^{\beta_1^j x^j} \\ \dots \\ e^{\beta_{k-1}^j x^j} \end{bmatrix}}{\sum_{j=0}^{k-1} e^{\beta_j^j x^j}} \end{array} \right. \quad (2)$$

In Eq. (2), x refers to the vector value. $\alpha()$ refers to a dimensional vector. e refers to the natural base. k refers to the maximum dimension value. j refers to the current dimension. β refers to an undetermined parameter. $f()$ refers to the probability function. T refers to the temperature coefficient used to adjust curve smoothness. Eq. (3) is the cross entropy of cost function.

$$H(p, q) = -\sum_{i=1}^N p(x_i) \log q(x_i) \quad (3)$$

In Eq. (3), H is the cross entropy of the cost function. p refers to the true distribution probability. q refers to the fitted distribution probability. x refers to the sample space. N is represented as a variable coefficient. Cross entropy mainly expresses the information quantity fully used to eliminate uncertainty. Then, the gradient descent algorithm was used to search for various parameters that optimize cost function. Finally, Backpropagation is used to transmit the optimized error information to model's initial neurons. So it can correct whole model's error to improve its accuracy. The model's excessive reliance on its own training data to fit iterative loop transmission can lead to a decrease in prediction accuracy. At this point, it is necessary to use regularization methods that modify the penalty term of cost function and regularization discarding methods that discard neurons to reduce model complexity and prevent overfitting in Equation (4).

$$\left\{ \begin{array}{l} L1 = H(p, q) + \alpha \sum_{i=1} |\beta_i| \\ L2 = H(p, q) + \alpha \sum_{i=1} \beta_i^2 \end{array} \right. \quad (4)$$

In Eq. (4), $L1$ is expressed as regularization that minimizes weight's absolute value. $L2$ is expressed as regularization that minimizes the square of weights. α is expressed as a regularization coefficient. β is expressed as a weight coefficient. $L1$ is suitable for avoiding overfitting by reducing weight density to simplify data while accurately modeling. $L2$ is suitable for avoiding overfitting in computer image recognition by attenuating weights. So combining the two can achieve good regularization results. When RNN is applied to NIDS to detect real data, it can effectively improve the accuracy of analyzing common attack data.

B. Design and Implementation of Domain Name Classification Network Model

The model needs to introduce DGA and CNN to address rare attacks such as malicious domain names [20, 21]. Compared to traditional network models, CNN can further

simulate the multi-layer NN architecture of brain to analyze complex information. It can improve data model's ability to analyze and judge complex feature relationships between network packets. Fig. 4 show its network structure.

According to Fig. 4, CNN contains a 5-layer structure. Among them, input layer can standardize multidimensional data to improve model's learning efficiency and result expression. Convolutional layer mainly uses excitation functions to assist in feature recognition and extraction of standardized input data. Pooling layer uses pooling functions to select the characteristics of the extracted feature map and screen key information, so as to reduce parameters number and realize feature data invariance to reduce subsequent calculation. Fully connected layer uses classification algorithms to perform nonlinear combination of transformed extracted features to achieve the function of a "classifier". The output layer uses logic or normalization functions to process data and output a numerical matrix. Eq. (5) and (6) represent the calculation of fully connected layers.

$$y_i = \text{relu}(w_i * y_{i-1} + b_i) \quad (5)$$

In Eq. (5), y_i is expressed as the output of i -th layer. w_i is expressed as a weight coefficient. b_i is expressed as an offset parameter. relu is expressed as activation function.

$$F(y) = i|x; \theta(w, b)) = e^{\theta(w, b)x} / \sum_{j=1}^k e_j^{\theta(w, b)x} \quad (6)$$

In Eq. (6), $F()$ is the classification function. y refers to the predicted project category value. x refers to the sample value. $\theta(w, b)$ refers to the classification parameter. e is represented as a natural base. k refers to the label type of classified data. The preprocessed data enters the output layer for normalization processing and outputs the result, which is mathematically expressed as Eq. (7).

$$T_{i,j} = (T_{i,j} - \min(T_{i,j})) / (\max(T_{i,j}) - \min(T_{i,j})) \quad (7)$$

In Eq. (7), $T_{i,j}$ is expressed as the characteristic values of a certain row and column. This model introduces DGA on the foundation of CNN, and collects samples from all network DSs and classifies them into small DSs according to certain standards for fine processing. Thus, an N-gram combined Character Based Deep Network (NCBDN) on the foundation of DGA was proposed. Fig. 5 shows the model process.

From Fig. 5, the model mainly consists of several parts, including data collection, data feature extraction and classification, abnormal data classification, and response strategy. The data detection model divides the overall network DS into smaller DS according to the network protocol, such as transmission control protocol, User Datagram Protocol and Internet control message protocol [22]. After further dividing each DS into normal and abnormal data, the experiment further classifies the abnormal DS into different attack types to form policy responses. Fig. 6 shows the data detection module and abnormal data learning module.

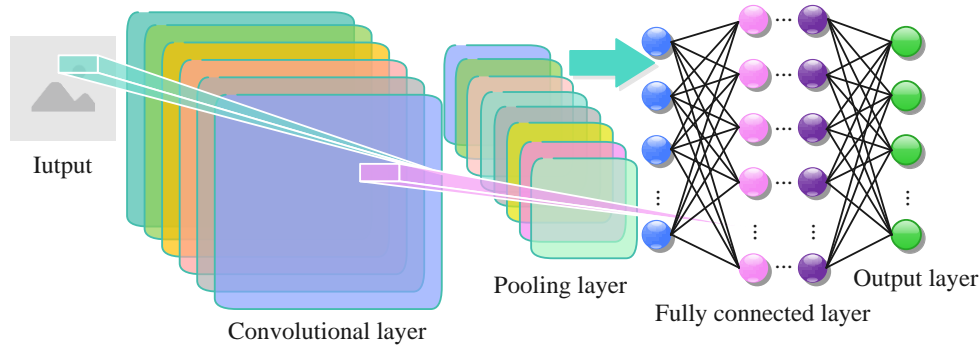


Fig. 4. CNN structure diagram.

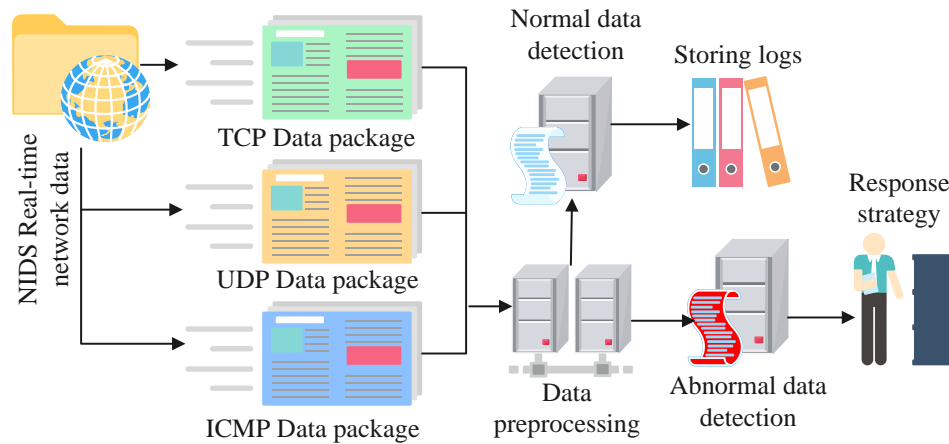


Fig. 5. Flowchart of NCBDN data probe model.

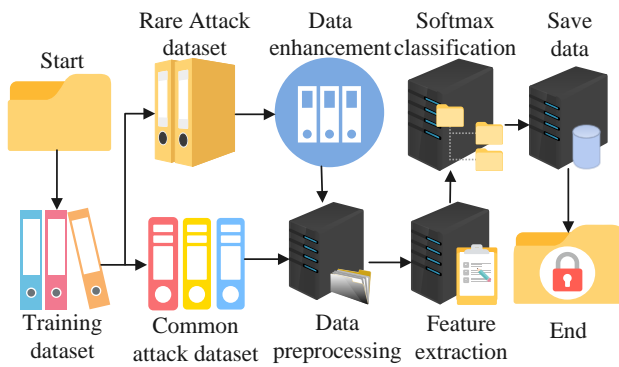


Fig. 6. Flowchart of data detection module and flow chart of abnormal data training module.

From Fig. 6, the model categorizes attack types into ordinary attacks and rare attacks. In order to form the initial training experience of the model, it is necessary to use the training set for initial training to form a basic feature library. To address malicious domain name attacks, it is necessary to first calculate the correlation coefficients between malicious domain names and normal domain names based on the characteristics of the DGA algorithm. Eq. (8) is its mathematical expression.

$$p = \frac{\sum_{i=1}^N (z_i - \bar{z})(e_i - \bar{e})}{\sqrt{\sum_{i=1}^N (z_i - \bar{z})^2} \sqrt{\sum_{i=1}^N (e_i - \bar{e})^2}} \quad (8)$$

In Eq. (8), p refers to the correlation coefficient. N refers to the total sample points. z_i and e_i respectively represent the spatial distribution values of individual normal and malicious domain names. \bar{z} and \bar{e} both represent the average domain name distribution value. According to the calculated correlation coefficient, it is necessary to perform data transformation on the incoming raw training data. This facilitates a more comprehensive analysis and classification of model in the next step. The preprocessing of data transformation first requires feature extraction of data gain. Eq.(9) is the definition of entropy.

$$Info(S) = -\sum_{i=1}^{\infty} Q_i \log(Q_i) \quad (9)$$

In Eq. (9), $Info$ is expressed as an information return function. S is expressed as a variable. Q is expressed as the probability of each variable value. So Equation (10) is the information required for identifying variables.

$$Info(L, S) = \sum_{i=1}^m |S_i|/|S| * Info(S_i) \quad (10)$$

In Eq. (10), L is expressed as a non-classified label value, and m is expressed as sample number. Combining Eq. (9) and (10) can ultimately obtain the required data gain in Eq. (11).

$$Gain(L, S) = Info(S) - Info(L, S) \quad (11)$$

In Eq. (11), $Gain$ is expressed as a gain function. After the preprocessed DS enters detection module, if the feature code matches the regular data feature code, it will be classified and archived. If the signature matches the abnormal feature, a policy response will be triggered. In addition, the judgment criteria for classifiers include multiple aspects. Eq. (12) refers to accuracy and false positive rate.

$$\begin{cases} Accuracy = (TP + TN) / (TP + TN + FP + FN) \\ FalsePositiveRate = FP / (FP + TN) \end{cases} \quad (12)$$

In Eq. (12), TP is true class, TN is true negative class, FP is false positive class, and FN is false negative class. The mathematical expressions for precision ($Precision$) and recall ($Recall$) in Eq. (13) are as follows.

$$\begin{cases} Precision = TP / (TP + FP) \\ Recall = TP / (TP + FN) \end{cases} \quad (13)$$

Eq. (14) is the mathematical expression for micro precision ($microP$) and micro recall ($microR$).

$$\begin{cases} microP = \overline{TP} / (\overline{TP} + \overline{FP}) \\ microR = \overline{TP} / (\overline{TP} + \overline{FN}) \end{cases} \quad (14)$$

The micro F1 equation in Eq. (15) can be obtained from Equation (14).

$$microF1 = (2 \times microP \times microR) / (microP + microR) \quad (15)$$

In Equation (15), $microF1$ refers to average micro harmonic value, and Equation (16) refers to macro F1.

$$macroF1 = ((2/n^2) \sum_1^n P_i R_i) / ((1/n) \sum_1^n P_i + (1/n) \sum_1^n R_i) \quad (16)$$

In Eq. (16), $macroF1$ is average macro harmonic value, P is precision, and R is recall. The intelligent judgment accuracy can be comprehensively evaluated through classifier's judgment criteria. This facilitates timely adjustment and optimization of system parameters to ensure stable model operation within the optimal range for a long time.

IV. MODEL VALIDATION AND DATA ANALYSIS

Based on the above algorithm analysis, NCBDN has higher precision in detecting, analyzing and classifying rare attacks, compared with the traditional detection model including malicious domain names based on DGA algorithm. To verify model performance advantage in detecting and identifying malicious domain names generated by DGA, NCBDN was used in this experiment to compare data detection classification of 16 domain names generated by DGA. Fig. 7 shows the

comparison of four testing standards for binary (B) character detection.

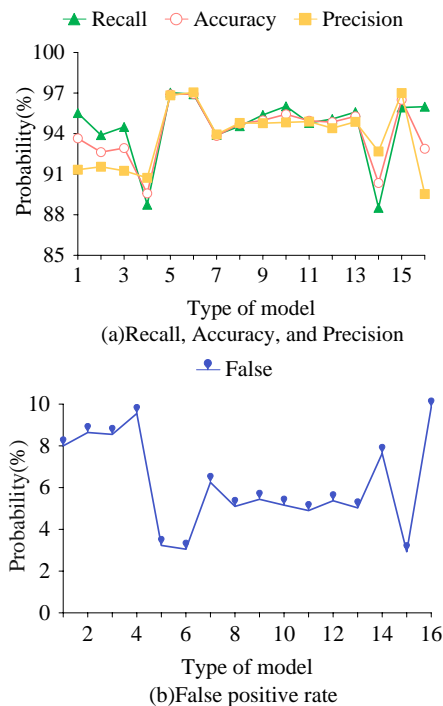


Fig. 7. Standard probability plot for bigram character detection.

From Fig. 7, NCBDN based on binary syntax has an average accuracy of about 94%, an average detection rate of about 94%, an average precision of about 93%, and an average error rate of about 6% for the recognition of 16 malicious domain names generated by DGA algorithm. Fig. 8 shows the comparison of trigram (T) character detection indicators.

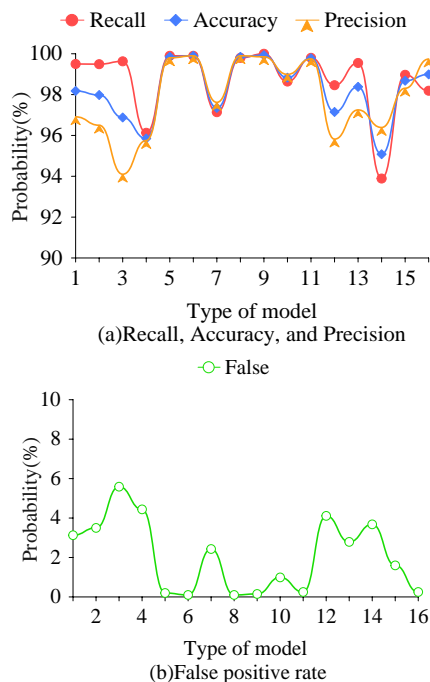


Fig. 8. Standard probability plot for trigram character detection.

From Fig. 8, NCBDN recognition rate with ternary syntax for malicious domain names is further improved by about 98%, detection rate by about 98%, precision by about 98%, and error rate by about 2%. So when N=3, NCBDN has the highest efficiency in identifying malicious domain names. To further validate model advantages in identifying malicious domain names, NCBDN binary and ternary models were compared with six models for malicious domain name recognition. They include Logistic Regression (LR), Naive Bayesian Model (NB), K-Nearest Neighbor (KN), and Support Vector Machine (SV) in Fig. 9.

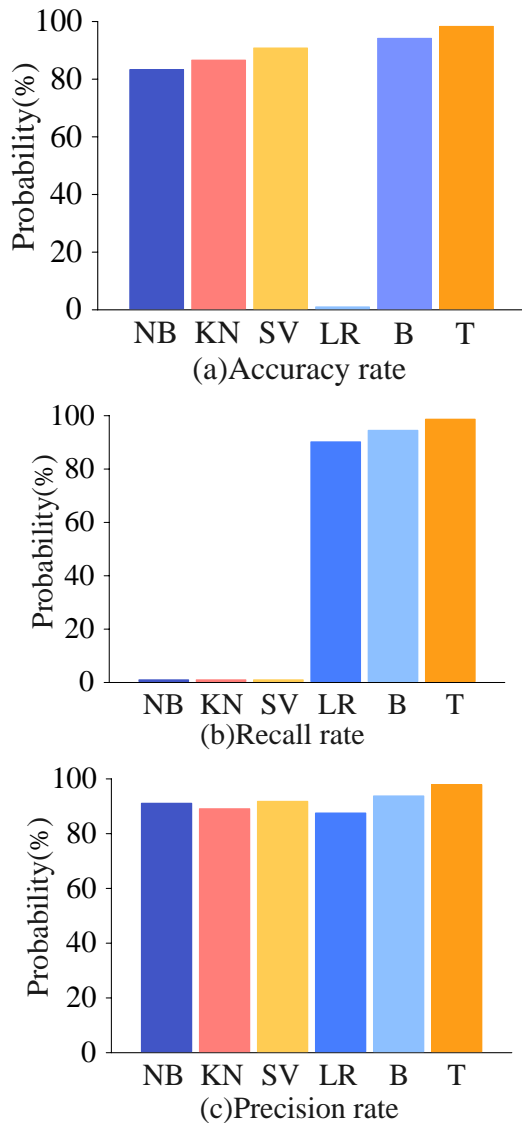


Fig. 9. Comparison of malicious domain name recognition probabilities of 17 models.

In Fig. 9, only NCBDN model under ternary syntax has an average DR of over 98%. This model can more effectively identify malicious domain names generated by DGA. This model adopts feature extraction methods to enhance the recognition probability of this model against unknown attacks. So seven methods were used, including Empirical Mode Decomposition (EMD), Hierarchical Spatial temporal features

based ID System (HAST-IDS), Text-CNN and RF based ID System (TR-IDS), Analog Network ID System (A-NIDS), Recursive Feature Addition(RFA), SVM and Intuitionistic Fuzzy Set for Anomaly Detection (IFSE-AD). The training datasets are Information Security Center of Excellence 2012 (ISCX2012) and Canadian Institute for Cybersecurity ID System 2017 (CICIDS2017).

The ISCX2012 dataset was collected by the Network Security Laboratory of Dalhousie University in Canada to provide a real network traffic dataset for research in network intrusion detection and traffic analysis. The ISCX2012 dataset consists of two sub-datasets. The Botnet-based sub-dataset contains normal traffic from the real network and malicious traffic from Zeus and ZeroAccess, among which normal traffic accounts for about 85% of the whole dataset and malicious traffic accounts for about 15%. The DDOS-based sub-dataset contains normal traffic from the real network and malicious traffic from different types of DDOS attacks, where normal traffic accounts for about 80% of the whole dataset and malicious traffic accounts for about 20%.

The CICIDS2017 dataset was collected in collaboration with the Institute for National Security and Counterterrorism (INSA) laboratory and the Cybersecurity Laboratory at Concordia University. This paper aims to provide a diverse network intrusion detection dataset for evaluating and comparing different intrusion detection systems. The CICIDS2017 dataset consists of several sub-datasets, the most commonly used of which are: The Botnet-based sub-dataset contained normal traffic from the real network and malicious traffic from malicious botnets such as Mirai, Gafgyt and TBot, among which normal traffic accounted for about 60% of the whole dataset and malicious traffic accounted for about 40%. The DDOS-based sub-dataset contains normal traffic from the real network and malicious traffic from different types of DDOS attacks, where normal traffic accounts for about 80% of the whole dataset and malicious traffic accounts for about 20%.

It is important to note that the components and proportions of the dataset may vary from version to version and use.

Table I shows the data results and the recognition efficiency of the model for unknown attacks.

In Table I, N/A indicates that the return value is invalid. From Table I, as sample number increases, model recognition accuracy increases. The recognition accuracy of EMD is 90% when data volume reaches around 10000. HAST-IDS has a recognition accuracy of over 90% with a data volume of around 900000. The accuracy of TR-IDS exceeds 90% when data volume reaches 30000. The accuracy of A-NIDS is close to 90% when data volume reaches 80000. The recognition rate of RFA is about 70% when data volume is 30. The recognition probability of IFSE-AD is over 95% when data volume reaches 20000. The accuracy of NCBDN can reach over 95% when data volume is 20. So this model has certain advantages in learning efficiency. The model has a high recognition probability for ordinary attacks, but the recognition probability for rare attacks is unknown. Fig. 10 shows the identification probability for verifying against rare attacks.

TABLE I. LEARNING AND RECOGNITION EFFICIENCY OF DIFFERENT PROBE MODELS

Method	Data Set	Sample Size	A (%)	R (%)
EMD	ISCX2012	9548	91.20	91.78
HAST-IDS	ISCX2012	819167	89.46	85.69
HAST-IDS	ISCX2012	908734	98.92	95.10
TR-IDS	ISCX2012	31407	91.45	91.76
A-NIDS	ISCX2012	80645	88.10	N/A
RFA	ISCX2012	24	76.50	70.40
RFA	ISCX2012	488	91.54	88.47
SVM	CICIDS2017	N/A	N/A	93.89
IFSE-AD	CICIDS2017	2422	92.55	N/A
IFSE-AD	CICIDS2017	23194	96.79	N/A
NCBDN	ISCX2012	8	96.26	97.99
NCBDN	ISCX2012	16	98.73	98.91
NCBDN	CICIDS2017	8	93.58	99.24
NCBDN	CICIDS2017	16	96.91	99.55

In Fig. 10, five types of data are introduced, namely Denial of Service (Dos), User to Root (U2R), Remote to Local (R2L), Probe, and Normal. U2R and R2L are rare attacks. As meta grammar increases from primeval number to 4, their recognition probability against rare attacks increases first and then decreases. When N=3, its recognition probability is the highest, about 85% for U2R and 92% for R2L. To further confirm model recognition performance when N=3, the detection accuracy was compared with Hierarchical ID model (HIDM), Managed ID model (MIDM). Fully connected detection model (FCDM), and CNN model (CNNM) in Fig. 11, respectively.

In Fig. 11, NCBDN has significant advantages over traditional models in terms of learning efficiency and detection efficiency of rare attacks. NCBDN has a high learning efficiency for unknown attacks and a high probability of identifying rare attack data.

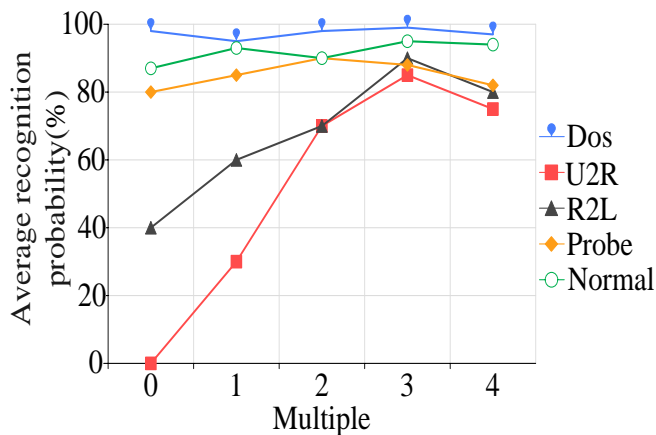


Fig. 10. Comparison plot of data augmentation probabilities.

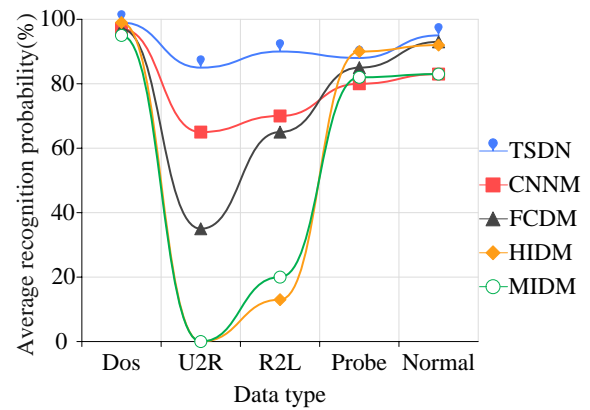


Fig. 11. Comparison of common detection models.

V. DISCUSSION

This study aims to explore the network intrusion detection method based on deep learning technology. With the rapid development of computer networks, cyberspace carries more and more high-value information, and the problem of network security is becoming more and more serious. The traditional intrusion detection system has the problems of slow detection speed and high false alarm rate, so it is necessary to use deep learning technology to improve the detection effect.

Several deep learning-based network intrusion detection methods have been proposed in the research. Firstly, by combining recurrent neural network with Gated Recurrent Unit (GRU) and Multilayer Perceptron (MLP), a network intrusion detection method based on GRU was proposed. Experimental results show that the proposed method has higher detection rate and lower false alarm rate. Secondly, a domain name detection method based on semantic expression is proposed. The detection of malicious domain names is realized by using deep convolutional neural networks. Experimental results show that the method has a good detection effect on domain names generated by different types of algorithms.

In addition, a network intrusion detection method for small sample based on meta-learning framework is proposed. This method realizes the detection of network intrusion behaviors in small sample scenarios through differential expression. Experimental results show that the proposed method has higher detection rate in small sample scenarios.

Finally, a low-rate denial of service attack detection method based on hybrid deep neural network was proposed. This method realizes the detection of low-rate DOS attacks by one-dimensional convolutional neural network and gated recurrent unit. Experimental results show that the proposed method has good detection effect in both large sample and small sample scenes.

In summary, the research has proposed a variety of effective network intrusion detection methods by applying deep learning techniques. Experimental results show that these methods can achieve good detection results in different scenes. Future work can further study the application of deep learning technology in the field of network security, solve the security problem of deep neural network itself, and further improve the

intrusion detection method according to the characteristics of the development of emerging networks.

VI. CONCLUSION

In response to issues such as low learning and recognition efficiency in malicious domain name attacks, this study first based on traditional NIDS models and introduced a recurrent neural deep learning network model containing GRU to make it more intelligent. To enable it to identify unknown rare attacks more quickly and accurately, this research further deepened and constructed NCBDN based on DGA algorithm. The experiment trained and learned the model using DGA algorithm and a dataset containing attack DS such as Dos, U2R, R2L, and Probe, as well as normal DS. The optimal DGA, Dos, 99%, U2R, 85%, R2L, Probe, 81%, and Normal recognition probabilities for NCBDN for malicious domain names and five types of data were 98%, 97%, and 97%, respectively. The traditional ID system has a low comprehensive average recognition rate for DGA and two rare attacks, with HIDM being 0% and 16%, and MIDM being 0% and 20%, respectively. The average recognition rate of general NN is relatively high, with FCDM being 35% and 65%, and RNNM being 68% and 71%, respectively. The highest recognition rates for NCBDN are 98% and 92%, respectively. NCBDN model is obtained by improving N-gram of the traditional network based on DGA. This model not only ensures high recognition probability for normal data and ordinary attack data. And it further improves learning efficiency for unknown attacks and the recognition probability for rare attack data. However, NCBDN has low efficiency in collecting threat data in large network traffic. Therefore, model detection efficiency for threats in big data needs to be further improved.

REFERENCES

- [1] K. Tsiknas, D. Taketzis, K. Demertzis, and C. Skianis, "Cyber threats to industrial IoT: A survey on attacks and countermeasures", *IoT*, vol. 2, pp. 163-186, January 2021.
- [2] H. Guo, J. Li, J. Liu, N. Tian, and N. Kato, "A survey on space-air-ground-sea integrated network security in 6G", *IEEE Commun. Surv. Tutorials*, vol. 24, pp. 53-87, November 2021.
- [3] M. H. U. Sharif and M. A. Mohammed, "A literature review of financial losses statistics for cyber security and future trend", *World J. Adv. Res. Rev.*, vol. 15, pp. 138-156, August 2022.
- [4] W. Duo, M. C. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges", *IEEE/CAA J. Automatica Sin.*, vol. 9, pp. 784-800, April 2022.
- [5] P. Kumar, G. P. Gupta, and R. Tripathi, "Design of anomaly-based intrusion detection system using fog computing for IoT network", *Automatic Contr. Comput. Sci.*, vol. 55, pp. 137-147, May 2021.
- [6] X. Zhou, W. Liang, W. Li, K. Yan, S. Shimizu, I. Kevin, and, K. Wang, "Hierarchical adversarial attacks against graph-neural-network-based

- IoT network intrusion detection system", *IEEE Internet Things J.*, vol. 9, pp. 9310-9319, November 2021.
- [7] P. Nimbalkar and D. Kshirsagar, "Feature selection for intrusion detection system in Internet-of-Things (IoT)", *ICT Express*, vol. 7, pp. 177-181, June 2021.
- [8] A. Moubayed, L. Yang, and A. Shami, "MTH-IDS: A multitiered hybrid intrusion detection system for internet of vehicles", *IEEE Internet Things J.*, vol. 9, pp. 616-632, May 2021.
- [9] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam detection using bidirectional transformers and machine learning classifier algorithms", *J. Comput. Cogn. Eng.*, vol. 2, pp. 5-9, April 2023.
- [10] I. Hidayat, M. Z. Ali, and A. Arshad, "Machine learning-based intrusion detection system: An experimental comparison", *J. Comput. Cogn. Eng.*, vol. 2, pp. 88-97, July 2022.
- [11] A. Binbusayyis and T. Vaiyapuri, "Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM", *Appl. Intell.*, vol. 51, pp. 7094-7108, February 2021.
- [12] N. Wang, Y. Chen, Y. Xiao, Y. Hu, W. Lou, and Y. T. Hou, "Manda: On adversarial example detection for network intrusion detection system", *IEEE Trans. Dependable Secur. Comput.*, vol. 20, pp. 1139-1153, February 2022.
- [13] S. Krishnaveni, S. Sivamohan, S. S. Sridhar, and S. Prabarakan, "Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing", *Cluster Comput.*, vol. 24, pp. 1761-1779, January 2021.
- [14] A. H. Azizan, S. A. Mostafa, A. Mustapha, C. F. M. Foozy, M. H. A. Wahab, M. A. Mohammed, and B. A. Khalaf, "A machine learning approach for improving the performance of network intrusion detection systems", *Ann. Emerg. Technol. Comput. (AETIC)*, vol. 5, pp. 201-208, March 2021.
- [15] O. Almomani, "A hybrid model using bio-inspired metaheuristic algorithms for network intrusion detection system", *Comput., Mater. Contin.*, vol. 68, pp. 409-429, March 2021.
- [16] M. Rashid, J. Kamruzzaman, T. Imam, S. Wibowo, and S. Gordon, "A tree-based stacking ensemble technique with feature selection for network intrusion detection", *Appl. Intell.*, vol. 52, pp. 9768-9781, January 2022.
- [17] M. U. Ilyas and S. A. Alharbi, "Machine learning approaches to network intrusion detection for contemporary internet traffic", *Comput.*, vol. 104, pp. 1061-1076, January 2022.
- [18] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. R. Müller, "Explaining deep neural networks and beyond: A review of methods and applications", *Proc. IEEE*, vol. 109, pp. 247-278, March 2021.
- [19] W. Zhang, H. Li, L. Tang, X. Gu, L. Wang, and L. Wang, "Displacement prediction of Jiuxianping landslide using gated recurrent unit (GRU) networks", *Acta Geotech.*, vol. 17, pp. 1367-1382, April 2022.
- [20] M. Tripathi, "Analysis of convolutional neural network based image classification techniques", *J. Innov. Image Process. (JIIP)*, vol. 3, pp. 100-117, June 2021.
- [21] V. Ravi, M. Alazab, S. Srinivasan, A. Arunachalam, and K. P. Soman, "Adversarial defense: DGA-based botnets and DNS homographs detection through integrated deep learning", *IEEE Trans. Eng. Manage.*, vol. 70, pp. 249-266, March 2021.
- [22] M. Aboubakar, M. Kellil, and P. Roux, "A review of IoT network management: Current status and perspectives", *J. King Saud University-Computer Inform. Sci.*, vol. 34, pp. 4163-4176, July 2022.

Dynamic Modelling of Hand Grasping and Wrist Exoskeleton: An EMG-based Approach

Mohd Safirin Bin Karis¹, Hyreil Anuar Bin Kasdirin², Norafizah Binti Abas³, Muhammad Noorazlan Shah Bin Zainudin⁴, Sufri Bin Muhammad⁵, Mior Muhammad Nazmi Firdaus Bin Mior Fadzil⁶

Department of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia¹

Department of Electrical Engineering, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia^{2,3,6}

Department of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka, Malacca, Malaysia⁴

Department of Software Engineering and Information System-Faculty of Science and Information Technology, Universiti Putra Malaysia Melaka, Serdang, Selangor, Malaysia⁵

Abstract—Human motion intention plays an important role in designing an exoskeleton hand wrist control for post-stroke survivors especially for hand grasping movement. The challenges occurred as sEMG signal frequently being affected by noises from its surroundings. To overcome these issues, this paper aims to establish the relationship between sEMG signal with wrist angle and handgrip force. ANN and ANFIS were two approaches that have been used to design dynamic modelling for hand grasping of wrist movement at different MVC levels. Input sEMG signals value from FDS and EDC muscles were used to predict the hand grip force as a representation of output signal. From the experimental results, sEMG MVC signal level was directly proportional to the hand grip force production while hand grip force signal values will depend on the position of wrist angle. It's also concluded that the hand grip force signal production is higher while the wrist at flexion position compared to extension. A strong relationship between sEMG signal and wrist angle improved the estimation of hand grip force result thus improved the myoelectronic control device for exoskeleton hand. Moreover, ANN managed to improve the estimation accuracy result provided by ANFIS by 0.22% summation of integral absolute error value with similar testing dataset from the experiment.

Keywords—Hand grasping; wrist control; ANN; ANFIS; exoskeleton wrist design

I. INTRODUCTION

Dexterous human hand movement completed the routine of human daily activities by providing specific hand gestures for task movements, such as object grasping and posture maintenance [1],[2]. Recognizing that hand movement is changeable, as human grasping requires varying grip force and wrist angles to execute various activities [3]. Moreover, hand movement activities induced by user motion intention involves muscle contractions which can be monitored using surface electromyography (sEMG) data, resulting in force output [4],[5],[6]. However, the variation in wrist angle position associated with results in variation of sEMG signal amplitude, which would have a significant impact on the accuracy of grasp force estimation [7].

However, several diseases, including stroke, can have a negative impact on human hand function. Undeniably stroke is a major public health issue in many countries [8],[9]. In 2020, the World Health Organization (WHO) reported that 21,592

people in Malaysia had died from a stroke, accounting for 12.85% of all deaths in the country [10]. According to the 2013 Global Burden of Disease study, this disease currently ranks third among the most significant contributors to disability-adjusted life years [11], [12]. In general, approximately three-quarters of stroke survivors are still suffer from their post-stroke effects [13]. One of the most prevalent disabling effects of a stroke is upper limbs impairment [14]. Based on statistic values, between 55% to 75% of stroke survivors lost of their hand ability. This hand disability can make their survivors dependent on others for help with activities of daily living, which can lower their quality of life [15], [16]. In the sense of that, since 1952, many researchers have looked at the concept of surface electromyography (sEMG) signals production as a means of improving Human Machine Interaction (HMI) for the benefit of post-stroke survivors [17].

Providing a path to integrate HMI has always triggers a challenge among all the research. In such tasks or activities, the robot should be designed to match the human arm's dexterity and skill [18]. Therefore, it is crucial to examine the biomechanical model of human muscle force and transfer it to the robot control in order to establish smooth interaction between the human and the robot instead of simple stiff interaction [18]. Although it is difficult to estimate grip force from sEMG signals, successful force recognition can aid in the design of a usable interface for natural and accurate EMG-based robot control [19]. The estimation of a generated force from sEMG signals enables the control of robotic equipment such as exoskeletons or prostheses in real time applications [20], [21], [22].

Realizing the importance understanding of sEMG signals, wrist angle and force excitation in forming the hand movement activities, the exoskeleton hand has been created in Solidwork software and converted in visual Matlab 2017a environment to imitate the natural human hand movement. The input designed for exoskeleton hand was the sEMG signals has been analysed at different wrist angle position (flexion and extension) with different Maximum Voluntary Contraction (MVC) level of hand grasping. The output function of sEMG was the estimation of hand grip force as it was needed to improve the myoelectric control system performance [6].

Nicola Secciani et al. proposed a control strategy based on "classification loop" and a "actuation loop" to control the movement of exoskeleton hand for free grasp, spherical grasp, and cylindrical grasp [23]. In 2019, Jing Luo et. al., used Neural Network Based Approach to estimate the force based on received input EMG signals [18]. The research continues as He Mao et. al., and Jiaqi Xue et. al., used EMG signals to estimate force and angle that represent hand movement such as wrist flexion/extension, ulnar/radial deviation, pronation/supination, and grip in 2023 [3],[19]. By recognising the significance of predicting future output results based on EMG input, more opportunities of improvement can be realised, as this relationship can strengthen control area in exoskeleton hands and prostheses section for future development [24].

This paper aims to analyse the relationship between sEMG signals, wrist angle and hand grip force generation at different MVC level using Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) dynamic modelling system for exoskeleton hand. The output of this relationship has been expected to improve the understanding on hand grasping control system strategy and selecting the best dynamic modelling approach for exoskeleton hand.

II. RECENT DEVELOPMENT

EMG-force relationship always been a highlight in analysing the hand grasping process related with HMI. When using a tool, such as interacting with a robot that requires a high degree of dexterity, the dynamics of a person's arms can have a significant impact on that person's daily activities [18]. Moreover, the dynamic modelling develops for the system always come with an issue of finding the suitable approach to form the relationship between input of EMG signals generated from user motion intention based on muscle contraction towards force generated as an estimated output. Daniele et al. employed a multiple linear regressions strategy to reduce the reconstruction error of exerted endpoint forces from EMG force estimates [5]. Gelareh et al. said that other researchers frequently employ ANN and Support Vector Machines (SVM) to discover mappings processes between EMG and force [4]. Furthermore, Galareh et al. revealed that researchers employed system identification approaches such as polynomial estimation, linear regression, and fast orthogonal search (FOS) to estimate force from sEMG signals [4], [6]. Jiaqi et al. focused on feature design by comparing the performance of EMG linear envelope (ENV) and non-linear EMG to muscle activation mapping (ACT) to obtain optimal force estimate performance [19].

ANN are one of the methods that can be used to define a connection between an input with an output. It acts as a black box model to approximate a complex nonlinear mapping between the sEMG signals towards their wrist angle or force generated equivalent to the signal muscle contractions related to it. ANN can learn from observation of mixture muscle signals and did not require any understanding of biological phenomena of exoskeleton hand system such as mathematical equation to express the relationship between input and output. According to Changmok et al., ANN is computationally efficient and has been implemented in various real-time

systems [25]. Numerous similar research publications have demonstrated the effectiveness of neural networks in recognising EMG patterns [26]. Francisco et al., 2020, created a multiclass categorization model using a regression algorithm and neural networks to control an anthropomorphic robotic system with three degrees of freedom that can accurately remote the robot arm to specified positions in a state machine [27].

The neural-fuzzy-based myoelectric control system is another scheme for controlling an upper limb exoskeleton-type exoskeleton. Neural fuzzy is defined as the combination of a neural network and fuzzy logic in modern artificial intelligence theory [28]. Kazuo et al. pioneered the neuro-fuzzy myoelectric control system, in which fuzzy logic was comprised of "IF and THEN" statements and the fuzzy modifier was a fully connected neural network [29], [30]. The neural network must tune the fuzzy logic using the EMG signals. Typically, data-driven approaches for ANFIS network synthesis are based on clustering a training set of numerical samples of the unknown function to be approximated. Since then, ANFIS networks have been successfully applied to classification tasks, rule-based process controls and pattern recognition problems [31]. According to Jirui et al., ANFIS can also be used to represent an effective neural network strategy for solving function estimation problems [32]. Moreover, Song Yu et. al., managed to prove the result from ANFIS is better than Tonic Stretch Reflex Threshold (TSRT) approach used for elbow flexors or extensors in their research [33].

Both ANN and ANFIS were established mapping methods that can be used to design a dynamic modelling for exoskeleton hand system. However, to enhance the performance of myoelectric control strategies for exoskeleton hand system, the selection of mapping method to form a dynamic modelling needed to be carefully selected. According to Mao et al., intuitive control that mimics human hand movement as closely as feasible has been greatly praised [3]. When interacting with the external environment, humans typically regulate their force at different wrist angle positions to ensure good operation performance [18]. However, there has been little research into the simultaneous estimation of hand grip force and wrist angles in free space, which mimic the biological functions of human hands.

III. METHODOLOGY

A. Mechanical Hand Design

The exoskeleton hand was designed to mimic the natural human hand movement. From ten male subjects ages from 21 to 40 years old, all the anthropometric hand measurement were taken. One degrees of freedom (DoF) of the wrist angle position has been highlighted in this exoskeleton hand designed covered two types of gestures: hand grasping at -45° (flexion) and 45° (extension) showed in Fig. 1. Since the wrist exoskeleton hand can be moved to achieve wrist desired angle, it is completely actuated.

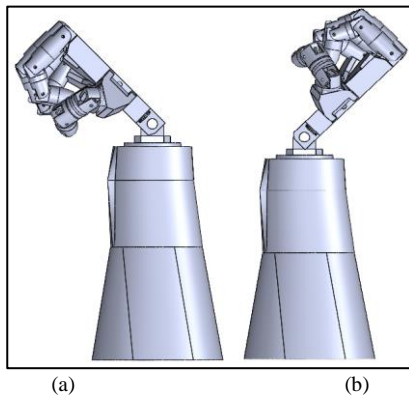


Fig. 1. Exoskeleton hand designed (a) flexion position (b) extension position.

B. EMG Data Collection

All experiments procedure were approved by the University Ethical Committee or Centre for Research and Innovation Management (CRIM) at University Technical Malaysia Melaka (UTeM) Malaysia. The experiment used a Hand Dynamometer, LabQuest Mini data acquisitions, Vernier EMG sensors, a personal computer with Logger Lite data-collection software, Stopwatch, Protector, and Kendall5400 diagnostic tab electrodes. Ten male subjects signed the researcher's consent form to undertake the hand grip pattern experiment at varying wrist angles at different MVC level. The experiment began after the subjects were fully briefed. Each experiment was repeated three times [34].

Flexor Digitorum Superficialis (FDS) and Extensor Digitorum Communis (EDC) EMG signal values have been employed in this research to represent each hand grasping wrist angle movement at different MVC level [3], [35], [23]. The medial epicondyle has been used to locate the muscles and the palpate scaphoid technique has been employed to determine the position of wrist movement [36], [37]. All subjects were in good health with non-neurological diseases and used their dominant hand for data collection.

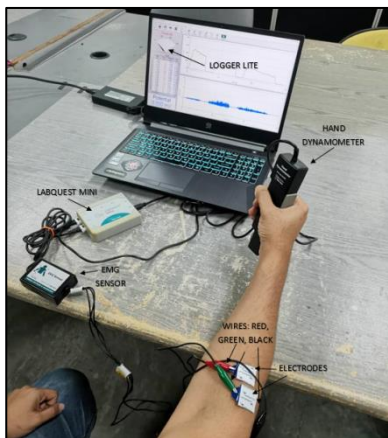


Fig. 2. Experimental set-up [38].

Fig. 2 illustrates how experimental procedures conducted. The maximum force (MVC) of the hand grasp is a measurement of the subject's strongest voluntary contraction. Electrode patches are put to the top of the abdominal muscles

of FDS and EDC. Samples were instructed to hold the hand dynamometer for five seconds at different hand grip strengths (20, 40, 60, 80, and 100% MVC level [34]. Each grip includes a two-second rest interval. The retrieved raw EMG signals were recorded using the Logger Lite programme. Fig. 3 shows the data collection for flexion and extension hand movement during the experiment.

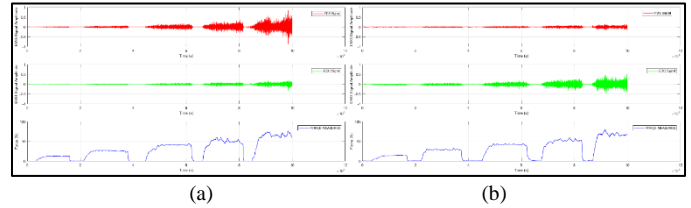


Fig. 3. Data collections for FDS, EDC and forces during (a) wrist angle at flexion (b) wrist angle at extension.

C. EMG Signal Processing

This paper adapted time-domain-based features using Waveform Length (WL) approach as it proven itself to be the best feature extraction method among RMS, MAV, IEMG and ZC as shown in Fig. 4 [39], [40], [41]. The sampling frequency was chosen at 1 kHz to suit the EMG signals range. The segmentation of input data was reduced at 50% analysis window increment. A second-order band-pass Butterworth filter was used for this experimental procedure [42]. The MVC method, which was uniquely recorded from each subject, has been used to standardize EMG measurement values. This approach scales the measurement value between 0 to 1 and most used normalization techniques in MVC-normalization [43], [44].

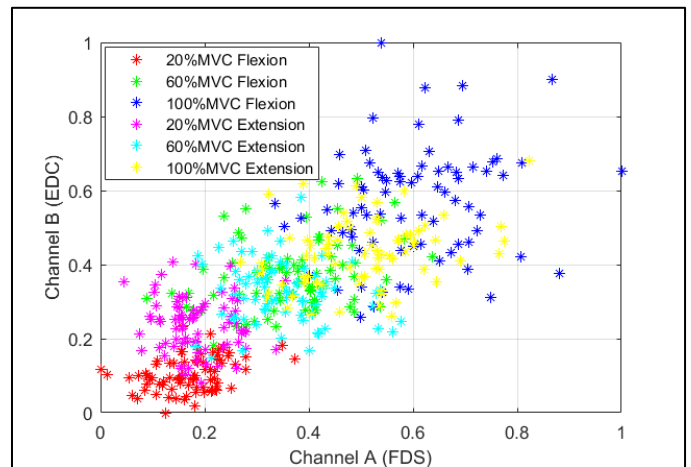


Fig. 4. WL feature extraction for 20%, 60% and 100% MVC at flexion and extension wrist angle.

D. Mapping Process

Modelling creates a connection between all usage parameters. It establishes a sequential connection between inputs and outputs. This modelling constructs an accurate transfer function to characterize system performance and the measured effectiveness of the selected modelling approach. Modelling, also known as mapping, is a data-based representation of numerous group design types. Since this paper recommended employing two mapping methods, ANN

and ANFIS were trained and evaluated using the same data set to ensure that their outputs were comparable.

1) *Method 1: Dynamic Modelling of Wrist Movement Using ANN:* ANN is one of the methods to generate a dynamic modelling for one's system. Depending on application complexity, neural networks can approximate nonlinear functions using adaptive weights on different layers [45]. From all the collected data set, two of them have been used to generate a training model for ANN approaches. The default setting has been used to generate this model representation as 70% dedicated for training, 15% for validation and 15% for testing. One number of hidden layers was used to connect two inputs with one output consisting of ten neurons shown in Fig. 4. The input were the EMG signals from FDS and EDC and output are the force generated from the hand grasping procedure at different wrist angles. Tangent sigmoid has been selected as ANN activation function and Levenberg-Marquardt has been selected as the training method [24]. The EMG data set taken from selected muscles has been arranged at 20%, 60%, 100% MVC at flexion state and 20%, 60%, 100% MVC at extension state to estimate the force generation at different wrist angle position. Fig. 5 shows architecture for ANN designed to estimate the hand grasping force.

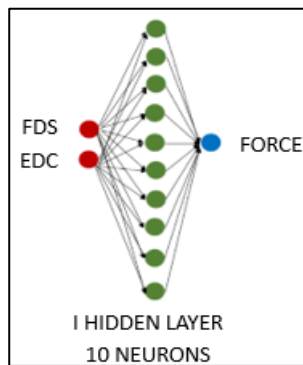


Fig. 5. ANN designed architecture.

2) *Method 2: Dynamic Modelling of Wrist Movement Using ANFIS:* ANFIS is another mapping method that can be used to create a dynamic modelling of a system. It has been built up from a combination of ANN and fuzzy logic to create its mapping block. It employs the fuzzification layer to map the input data to fuzzy sets using membership functions. The rule layer applies fuzzy if-then rules (Sugeno method) to capture the relationship between input variables and the output. The adaptation layer adjusts the parameters of the ANFIS model using a learning algorithm, allowing it to continuously improve its performance. For ANFIS setting, three sets of data coming from similar samples were used. two sets of them have been used to form a training block with 70% was dedicated for training, 15% for validation and 15% for checking. One hidden layer was chosen with ten neurons connected to the fuzzy rules to estimate the output value. FDS and EDC muscles sEMG signals have been chosen as an input while hand grasping force at different MVC levels has been selected as an output signal of a system. Number for membership function (mf) of ten neurons with combination of [2 8] and type of "gauss2mf" was set as an input and output

selected "constant" as their MF type. 20%, 60% and 100% MVC level for both flexion and extension of sEMG signal level have been arranged to predict the force hand grasping output. Fig. 6 shows an ANFIS designed architecture for the system.

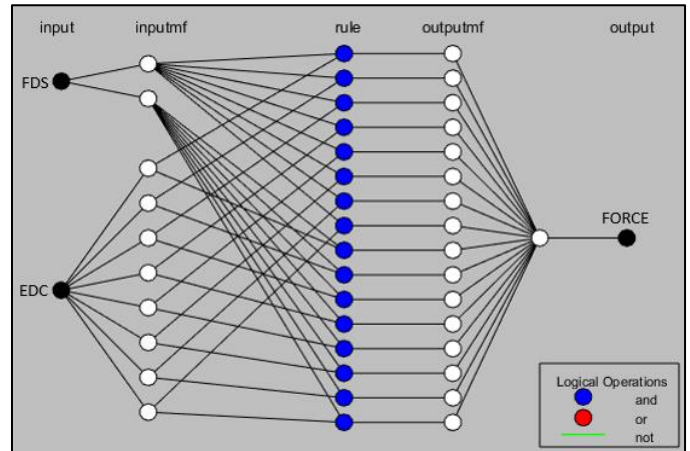


Fig. 6. ANFIS designed architecture.

IV. RESULTS

Fig. 7 depicts an analysis of hand grip forces. Within the same graph, a three-line force graph is plotted. Two of them are line graphs that depict the output from the ANN (magenta) and ANFIS (light blue) mapping processes, while the other (blue) was directly taken from the measurement process during the experimental procedure. As the wrist goes from flexion to extension, the graph is divided into two portions. The first section (wrist angle at flexion position) occurred between 0s and 233s. This section is organized into three subsections to represent the signal levels of 20% MVC, 60% MVC, and 100% MVC. The second part (wrist angle at extension position) existed between 234s and 467s. This section has also been separated into three subsections to represent the signal levels of 20% MVC, 60% MVC, and 100% MVC.

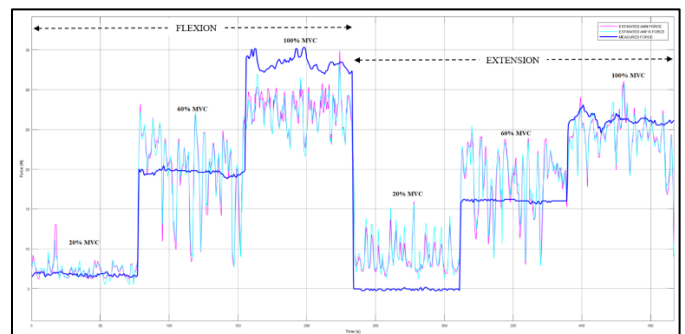


Fig. 7. Hand grip force analysis graph.

Fig. 8 depicts the absolute error for the ANN and ANFIS mapping methods. The magenta line graph is an error graph for the ANN approach, whereas the light blue graph is formed by the ANFIS approach. Both graph line plotting reveals a fluctuation pattern signal generated by both ways, as can be seen. Each method seems to have close estimated hand grip force and error values compare to each other and produce

quite a similar graph instead. These graphs were created by deducing the force measured value during the experiment process from the force estimation value obtained from both approaches. The total summation area under each graph representing the total summation error for each approaches used. According to Fig. 8, the sum of absolute error for the ANN technique was 19.33%, while the ANFIS approach was 19.55%. Based on this finding, ANN outperformed ANFIS in force estimation with a similar dataset and parameter settings.

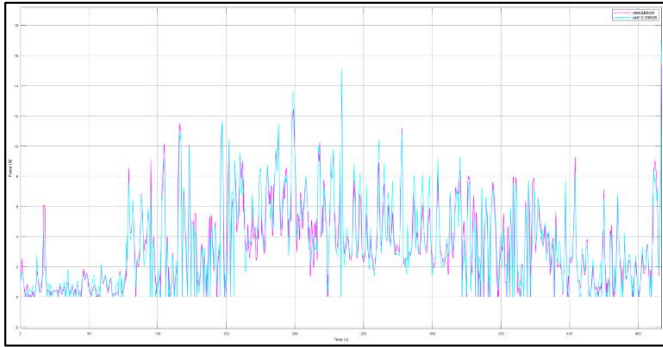


Fig. 8. Absolute error for ANN and ANFIS.

V. DISCUSSIONS

All of the hand grip forces MVC's plotting in the flexion part have higher values than their identical MVC's opponent in the extension section. The situation happened as referring to the normal human hand grasping movement. As the user motion intention directed the hand wrist angle to move towards human body, FDS muscle (flexion muscle) produce higher sEMG values compared to EDC muscle (extension muscle) signal which causing the flexion hand movement as shown in Fig. 3 [38]. For the flexion hand movement, the measured and estimated force values obviously demonstrated a higher signal level compared to their MVC in the other section. For the extension section, the hand wrist angle was pulled away from the human body, causing the EDC muscle to generate a greater sEMG value than the FCR muscle [3], [39]. As shown in Fig. 7, this extension movement degrades the hand grasp force value in each MVC level compared to the flexion movement.

At 20% MVC flexion section, the estimation graph plotting from both ANN and ANFIS mapping lingering closer to the measured value during the experiment. However, in the extension section, the estimation graph plotting introduced a small gap reading compared to the value from the experimental procedure. This could happen because the sEMG signal values from both muscles are almost the same when they are contracting at a lower level. The wrist angle position doesn't seem to have a big impact on how the signal number is read. Different case happened at 60% MVC level for both sections. Both muscles produce significant values of sEMG signals that causing the estimation force graph plotting for both approaches manage to differentiate between different wrist angle position. The fluctuation of force output graph might be coming from the nonlinearity of sEMG signals muscles contraction and the noise that interrupted during the data collection.

For 100% MVC, the force signals output estimation for the flexion section manages to have a higher value in their estimation but still does not give the same value as the measured one. The explanation for this is because subjects were instructed to flex their hands while grasping the hand dynamometer with 100% strength at 100% MVC of muscle contractions. Hand shaking commonly happened at this stage thus created a noisy environment while the data is being recorded, hence resulting an effect towards sEMG values used in the estimation process. For the extension section, at 100% MVC, the estimation graph for both hand grip force approaches achieves a nearly identical with the measured one due to a favourable environment for muscles contractions and a lower force measured value that allows the subject to perform well in hand grasping experimental procedure.

VI. CONCLUSION

Hand grasping is one of the most essential hand gestures for humans, including post-stroke patient survivors, to perform daily tasks. To comprehend and control the exoskeleton hand grasping gesture, the relationship between sEMG signal value, wrist angle, and hand grip force production triggered by the user's intention to grasp an object must be clearly understood. WL was chosen for feature extraction method in time domain in this study. To clarify the concept of force generation in both conditions, 20%, 60%, and 100% of the MVC level for flexion and extension wrist joint angle movement were analysed. As a consequence of the experiment procedure and dynamic modelling process, the sEMG MVC signal level was determined directly proportional to the generation of hand grip force. However, the hand grip force signal generated will depend on the wrist angle position. It was also determined that the hand grasp force signal production became greater when the wrist was in flexion as opposed to extension.

ANN and ANFIS were both dynamic modelling method used to analyse the hand grip force estimation. ANN has the capability to interpret unstructured data while, ANFIS used the strength from ANN and fuzzy to adapt with various environments to design a mapping system for exoskeleton hand. For the whole exoskeleton hand system, ANN and ANFIS needs a similar training and testing data set. Both approaches manage to generate its own dynamic model to represent the exoskeleton hand system. When compared to measured hand grip force recorded throughout the experiment process, ANN outperformed ANFIS by 0.22% absolute error with similar settings for both systems. When compared to ANFIS, the ANN technique produces a more accurate estimation of hand grip force output results.

The limitation of this study was the small number of neurons of ten provided for the ANN and ANFIS methods. Because this paper only focused on a similar number of neurons for output comparison, the value of neurons and their combination can be varied to produce a variety of possible output for the estimation results. Moreover, there are other regression method available they may need to be considered to improved estimation output results such as Support Vector Regression (SVR), Linear Regression (LR), Gaussian Process Regression (GPR), ensemble and decision tree.

ACKNOWLEDGMENT

The authors would like to thank the Ministry of Higher Education Malaysia for the financial support. The project is funded under Fundamental Research Grant Support. No (Ref: FRGS/1/2021/TK0/UTEM/02/54). The authors also want to thank Universiti Teknikal Malaysia Melaka for all the support.

REFERENCES

- [1] P. A. Banaszekiewicz and D. F. Kader, "Classic papers in orthopaedics," *Class. Pap. Orthop.*, pp. 1–624, 2014, doi: 10.1007/978-1-4471-5451-8.
- [2] T. Tsuji, P. G. Morasso, K. Goto, and K. Ito, "Human hand impedance characteristics during maintained posture," *Biol. Cybern.*, vol. 72, no. 6, pp. 475–485, 1995, doi: 10.1007/BF00199890.
- [3] H. Mao, Y. Zheng, C. Ma, K. Wu, G. Li, and P. Fang, "Simultaneous estimation of grip force and wrist angles by surface electromyography and acceleration signals," *Biomed. Signal Process. Control*, vol. 79, no. P1, p. 104088, 2023, doi: 10.1016/j.bspc.2022.104088.
- [4] G. Hajian, A. Etemad, and E. Morin, "Generalized EMG-based isometric contact force estimation using a deep learning approach," *Biomed. Signal Process. Control*, vol. 70, no. July, p. 103012, 2021, doi: 10.1016/j.bspc.2021.103012.
- [5] D. Borzelli, A. D'Avella, S. Gurgone, and L. Gastaldi, "Unconstrained and constrained estimation of a linear EMG-to-force mapping during isometric force generation," 2022 IEEE Int. Symp. Med. Meas. Appl. MeMeA 2022 - Conf. Proc., pp. 1–6, 2022, doi: 10.1109/MeMeA54994.2022.9856461.
- [6] N. Wang, K. Lao, X. Zhang, J. Lin, and X. Zhang, "The recognition of grasping force using LDA," *Biomed. Signal Process. Control*, vol. 47, pp. 393–400, 2019, doi: 10.1016/j.bspc.2018.06.011.
- [7] M. J. M. Hoozemans and J. H. Van Dieën, "Prediction of handgrip forces using surface EMG of forearm muscles," *J. Electromyogr. Kinesiol.*, vol. 15, no. 4, pp. 358–366, 2005, doi: 10.1016/j.jelekin.2004.09.001.
- [8] V. L. Feigin et al., "Global and regional burden of stroke during 1990–2010: Findings from the Global Burden of Disease Study 2010," *Lancet*, vol. 383, no. 9913, pp. 245–255, 2014, doi: 10.1016/S0140-6736(13)61953-4.
- [9] N. H. Omar, N. A. M. Nordin, C. S. Chui, and A. F. A. Aziz, "Functionality among stroke survivors with upper limb impairment attending community-based rehabilitation," *Med. J. Malaysia*, vol. 75, no. 2, pp. 146–151, 2020.
- [10] Zhou, Yang, and Wang, *World health statistics 2020: monitoring health for the SDGs, sustainable development goals*, vol. 21, no. 1, 2020.
- [11] V. L. Feigin et al., "Update on the global burden of ischemic and hemorrhagic stroke in 1990–2013: The GBD 2013 study," *Neuroepidemiology*, vol. 45, no. 3, pp. 161–176, 2015, doi: 10.1159/000441085.
- [12] C. J. L. Murray et al., "Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: Quantifying the epidemiological transition," *Lancet*, vol. 386, no. 10009, pp. 2145–2191, 2015, doi: 10.1016/S0140-6736(15)61340-X.
- [13] V. L. Feigin, B. Norrving, M. G. George, J. L. Foltz, G. A. Roth, and G. A. Mensah, "Prevention of stroke: A strategic global imperative," *Nat. Rev. Neurol.*, vol. 12, no. 9, pp. 501–512, 2016, doi: 10.1038/nrneurol.2016.107.
- [14] I. Faria-Fortini, S. M. Michaelsen, J. G. Cassiano, and L. F. Teixeira-Salmela, "Upper extremity function in stroke subjects: Relationships between the international classification of functioning, disability, and health domains," *J. Hand Ther.*, vol. 24, no. 3, pp. 257–265, 2011, doi: 10.1016/j.jht.2011.01.002.
- [15] C. De Diego, S. Puig, and X. Navarro, "A sensorimotor stimulation program for rehabilitation of chronic stroke patients," *Restor. Neurol. Neurosci.*, vol. 31, no. 4, pp. 361–371, 2013, doi: 10.3233/RNN-120250.
- [16] R. Teasell, M. J. Meyer, N. Foley, K. Salter, and D. Willems, "Stroke rehabilitation in Canada: A work in progress," *Top. Stroke Rehabil.*, vol. 16, no. 1, pp. 11–19, 2009, doi: 10.1310/tsr1601-11.
- [17] V. T. Inman, H. J. Ralston, J. B. De, M. B. Bertram Feinstein, and E. W. Wright, "Relation of human electromyogram to muscular tension," *Electroencephalogr. Clin. Neurophysiol.*, vol. 4, no. 2, pp. 187–194, 1952, doi: 10.1016/0013-4694(52)90008-4.
- [18] J. Luo, C. Liu, and C. Yang, "Estimation of EMG-Based force using a neural-network-based approach," *IEEE Access*, vol. 7, pp. 64856–64865, 2019, doi: 10.1109/ACCESS.2019.2917300.
- [19] J. Xue and K. W. C. Lai, "Dynamic gripping force estimation and reconstruction in EMG-based human-machine interaction," *Biomed. Signal Process. Control*, vol. 80, no. P1, p. 104216, 2023, doi: 10.1016/j.bspc.2022.104216.
- [20] P. Artemiadis, "EMG-based Robot Control Interfaces: Past, Present and Future," *Adv. Robot. Autom.*, vol. 01, no. 02, pp. 10–12, 2012, doi: 10.4172/2168-9695.1000e107.
- [21] R. M. Singh, S. Chatterji, and A. Kumar, "A review on surface EMG based control schemes of exoskeleton robot in stroke rehabilitation," *Proc. - 2013 Int. Conf. Mach. Intell. Res. Adv. ICMIRA 2013*, pp. 310–315, 2014, doi: 10.1109/ICMIRA.2013.65.
- [22] N. Parajuli et al., "Real-Time EMG Based Pattern Recognition Control Challenges and Future Implementation," *Sensors (Basel)*, vol. 19, no. 20, p. 4596, 2019.
- [23] N. Secciani, M. Bianchi, E. Meli, Y. Volpe, and A. Ridolfi, "A novel application of a surface ElectroMyoGraphy-based control strategy for a hand exoskeleton system: A single-case study," *Int. J. Adv. Robot. Syst.*, vol. 16, no. 1, pp. 1–13, 2019, doi: 10.1177/1729881419828197.
- [24] I. Chihi, L. Sidhom, and E. N. Kamavuoko, "Hammerstein–Wiener Multimodel Approach for Fast and Efficient Muscle Force Estimation from EMG Signals," *Biosensors*, vol. 12, no. 2, pp. 1–15, 2022, doi: 10.3390/bios12020117.
- [25] C. Choi, S. Kwon, W. Park, H. dong Lee, and J. Kim, "Real-time pinch force estimation by surface electromyography using an artificial neural network," *Med. Eng. Phys.*, vol. 32, no. 5, pp. 429–436, 2010, doi: 10.1016/j.medengphy.2010.04.004.
- [26] J. Huang, G. Li, H. Su, and Z. Li, "Development and Continuous Control of an Intelligent Upper-Limb Neuroprosthesis for Reach and Grasp Motions Using Biological Signals," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 52, no. 6, pp. 3431–3441, 2022, doi: 10.1109/TSMC.2021.3069084.
- [27] F. Pérez-Reynoso, N. Farrera, C. Capetillo, N. Méndez-Lozano, C. González-Gutiérrez, and E. López-Neri, "Pattern Recognition of EMG Signals by Machine Learning for the Control of a Manipulator Robot," *Sensors*, vol. 22, no. 9, 2022, doi: 10.3390/s22093424.
- [28] J. Vieira, F. Dias, and A. Mota, "Neuro-fuzzy systems: a survey," ... *Neural Networks Appl. Udine ...*, pp. 1–6, 2004, [Online]. Available: <http://dme.uma.pt/people/faculty/fernando.morgado/Down/483-343.pdf>
- [29] K. Kiguchi, T. Tanaka, and T. Fukuda, "Neuro-fuzzy control of a robotic exoskeleton with EMG signals," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 4, pp. 481–490, 2004, doi: 10.1109/TFUZZ.2004.832525.
- [30] K. Kiguchi, R. Esaki, and T. Fukuda, "Development of a wearable exoskeleton for daily forearm motion assist," *Adv. Robot.*, vol. 19, no. 7, pp. 751–771, 2005, doi: 10.1163/1568553054455086.
- [31] J. Addeh, A. Ebrahimzadeh, and H. Nazaryan, "A Research about Pattern Recognition of Control Chart Using Optimized ANFIS and Selected Features," *J. Eng. Technol.*, vol. 3, no. 1, p. 6, 2013, doi: 10.4103/0976-8580.107095.
- [32] J. Fu, R. Choudhury, S. M. Hosseini, R. Simpson, and J. H. Park, "Myoelectric Control Systems for Upper Limb Wearable Robotic Exoskeletons and Exosuits—A Systematic Review," *Sensors*, vol. 22, no. 21, pp. 1–31, 2022, doi: 10.3390/s22218134.
- [33] S. Yu, Y. Chen, Q. Cai, K. Ma, H. Zheng, and L. Xie, "A Novel Quantitative Spasticity Evaluation Method Based on Surface Electromyogram Signals and Adaptive Neuro Fuzzy Inference System," *Front. Neurosci.*, vol. 14, no. May, pp. 1–12, 2020, doi: 10.3389/fnins.2020.00462.
- [34] F. Xu, Y. Zheng, and X. Hu, "Estimation of joint kinematics and fingertip forces using motoneuron firing activities: A preliminary report," *Int. IEEE/EMBS Conf. Neural Eng. NER*, vol. 2021-May, pp. 1035–1038, 2021, doi: 10.1109/NER49283.2021.9441433.

- [35] J. M. Ochoa, D. G. Kamper, M. Listenberger, and S. W. Lee, "Use of an electromyographically driven hand orthosis for training after stroke," *IEEE Int. Conf. Rehabil. Robot.*, 2011, doi: 10.1109/ICORR.2011.5975382.
- [36] A. D. Sobel, K. N. Shah, and J. A. Katarincic, "The Imperative Nature of Physical Exam in Identifying Pediatric Scaphoid Fractures," *J. Pediatr.*, vol. 177, pp. 323-323.e1, 2016, doi: 10.1016/j.jpeds.2016.06.086.
- [37] V. Mendez, L. Pollina, F. Artoni, and S. Micera, "Deep learning with convolutional neural network for proportional control of finger movements from surface EMG recordings," *Int. IEEE/EMBS Conf. Neural Eng. NER*, vol. 2021-May, pp. 1074-1078, 2021, doi: 10.1109/NER49283.2021.9441095.
- [38] M. S. Karis, H. A. Kasdirin, N. Abas, W. H. M. Saad and M. S. M. Aras, "EMG BASED CONTROL OF WRIST EXOSKELETON," vol. 24, no. 2, pp. 391-406, 2023, <https://doi.org/10.31436/iiumej.v24i2.2804>.
- [39] N. Abas, W. M. Bukhari, M. A. Abas, and M. O. Tokhi, "Electromyography Assessment of Forearm Muscles: Towards the Control of Exoskeleton Hand," 2018 5th Int. Conf. Control. Decis. Inf. Technol. CoDIT 2018, pp. 822-828, 2018, doi: 10.1109/CoDIT.2018.8394906.
- [40] J. Lara, N. Paskaranandavadivel, and L. K. Cheng, "HD-EMG Electrode Count and Feature Selection Influence on Pattern-based Movement Classification Accuracy," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2020-July, pp. 4787-4790, 2020, doi: 10.1109/EMBC44109.2020.9175210.
- [41] A. Neacsu, J. Pesquet, and C. Burileanu, "ACCURACY-ROBUSTNESS TRADE-OFF FOR POSITIVELY WEIGHTED NEURAL NETWORKS Speech and Dialogue Laboratory University Politehnica of Bucharest , Bucharest , Romania Universit ´ e Paris-Saclay , CentraleSup ´ elec , Inria Centre de Vision Num ´," *ICASSP 2020 - 2020 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 8384-8388, 2020.
- [42] H. Hayashi, T. Shibanoki, and T. Tsuji, "A Neural Network Based on the Johnson SU Translation System and Related Application to Electromyogram Classification," *IEEE Access*, vol. 9, pp. 154304-154317, 2021, doi: 10.1109/ACCESS.2021.3126348.
- [43] A. Subasi and S. M. Qaisar, "Surface EMG signal classification using TQWT, Bagging and Boosting for hand movement recognition," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 7, pp. 3539-3554, 2022, doi: 10.1007/s12652-020-01980-6.
- [44] D. Copaci, D. Serrano, L. Moreno, and D. Blanco, "A high-level control algorithm based on sEMG signalling for an elbow joint SMA exoskeleton," *Sensors (Switzerland)*, vol. 18, no. 8, 2018, doi: 10.3390/s18082522.
- [45] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85-117, 2015, doi: 10.1016/j.neunet.2014.09.003.

Research on Semantic Segmentation Method of Remote Sensing Image Based on Self-supervised Learning

Wenbo Zhang, Achuan Wang

College of Computer and Control Engineering, Northeast Forestry University, Harbin, China

Abstract—To address the challenge of requiring a large amount of manually annotated data for semantic segmentation of remote sensing images using deep learning, a method based on self-supervised learning is proposed. Firstly, to simultaneously learn the global and local features of remote sensing images, a self-supervised learning network structure called TBSNet (Triple-Branch Self-supervised Network) is constructed. This network comprises an image transformation prediction branch, a global contrastive learning branch, and a local contrastive learning branch. The contrastive learning part of the network employs a novel data augmentation method to simulate positive pairs of the same remote sensing images under different weather conditions, enhancing the model's performance. Meanwhile, the model integrates channel attention and spatial attention mechanisms in the projection head structure of the global contrastive learning branch, and replaces a fully connected layer with a convolutional layer in the local contrastive learning branch, thus improving the model's feature extraction ability. Secondly, to mitigate the high computational cost during the pre-training phase, an algorithm optimization strategy is proposed using the TraIn method and sequential optimization theory, which increases the efficiency of pre-training. Lastly, by fine-tuning the model with a small amount of annotated data, effective semantic segmentation of remote sensing images is achieved even with limited annotated data. The experimental results indicate that with only 10% annotated data, the overall accuracy (OA) and recall of this model have improved by 4.60% and 4.88% respectively, compared to the traditional self-supervised model SimCLR (A Simple Framework for Contrastive Learning of Visual Representations). This provides significant application value for tasks such as semantic segmentation in remote sensing imagery and other computer vision domains.

Keywords—Computer vision; deep learning; self-supervised learning; remote sensing image; semantic segmentation

I. INTRODUCTION

With the rapid development of remote sensing satellite technology, remote sensing images are playing an increasingly critical role in various fields such as urban planning, resource exploration, and natural disaster prediction. Extracting useful information from the vast wealth of remote sensing geo-information has become a long-standing scientific challenge in remote sensing. Among the methods explored, semantic segmentation [2] has proven to be an effective approach.

In the field of semantic segmentation for remote sensing images, there are two main approaches: traditional methods

based on handcrafted feature descriptors and deep learning methods based on Convolutional Neural Networks (CNNs). Due to the complexity of background and scale differences in high-resolution remote sensing images, traditional methods have not been very effective. However, since Long et al. proposed the Fully Convolutional Neural Network (FCN) [3] in 2015, deep learning-based techniques for semantic segmentation in remote sensing images have made significant progress. This has led to the development of post-processing techniques based on probabilistic graphical models [4], global context modeling using multi-scale aggregations [5], and perpixel semantic modeling based on attention mechanisms [6].

For instance, Ronneberger et al. introduced the U-Net model [7], which employs an encoder-decoder architecture with lateral connections, enabling multi-scale recognition and feature fusion in the image. Similarly, Chen et al. proposed the DeepLabV3+ model [8], which utilizes a spatial pyramid structure to gather rich contextual information through pooling operations at various resolutions. Furthermore, it uses the encoder-decoder architecture to achieve precise object boundaries, thereby enhancing segmentation accuracy.

Despite the achievements made in deep learning-based semantic segmentation of remote sensing images in recent years [9][10], these methods all rely on large amounts of manually annotated data to train the neural network. This requirement not only consumes significant human resources but also reduces the efficiency of semantic segmentation. Therefore, the application of self-supervised learning [11] to semantic segmentation of remote sensing images has become a feasible method. Li et al. [12] proposed a multi-task self-supervised learning method for semantic segmentation of remote sensing images, which applied three pretext tasks [13] to self-supervised learning and achieved decent results. However, these pretext tasks only learn the global features of the image, lacking in the learning of local features of the image. Thus, how to effectively use these unannotated remote sensing data has become a major research focus in recent years.

The main contributions of this study are as follows:

1) To tackle the aforementioned challenges, a self-supervised semantic segmentation approach for remote sensing images is introduced, along with the design of a triple-branch self-supervised network named TBSNet. This network

uses an image transformation prediction branch and a global contrastive learning branch to learn global features of images, and a local contrastive learning branch to learn local features.

2) On this basis, the projection head structures in the global contrastive learning branch and the local contrastive learning branch are improved to enhance their performance. Specifically, in the projection head of the global contrastive learning branch, a combination of spatial attention mechanisms [14] and channel attention mechanisms [15] are used to better focus on the important parts of the feature map, thus improving the quality of feature representation. In the local contrastive learning branch, the original first fully connected layer is replaced with a convolutional layer, which enables the learning of richer local features.

3) Considering the heavy computational cost of pre-training, the TracIn method [16] and sequential optimization theory [17] are employed to optimize the model's pre-training process, reducing computational and time costs. Finally, the model is fine-tuned in the downstream task using a small amount of annotated data to achieve the expected semantic segmentation results.

The remaining structure of the article is outlined as follows: Section II introduces the background knowledge and relevant work. Section III describes the implementation details of the proposed method, including the network framework and optimization techniques. Section IV presents the experiments conducted and analyzes the obtained results. Section V provides the conclusions drawn from our experiments.

II. BACKGROUND

A. Self-supervised Learning

Self-supervised learning is a type of unsupervised learning [18], as shown in Fig. 1. Compared to supervised learning, it utilizes a large amount of unlabeled data through specially designed pretext tasks. This approach relies on pseudo-labels generated by the model itself, enabling it to learn high-level features from the input data. The model can then be further transferred to downstream tasks in actual applications. With a small amount of labeled data, the model can be fine-tuned to achieve, or even surpass, the performance of supervised learning. Generally, self-supervised learning can be divided into generative and contrastive categories [19] [20].

B. Pretext Task

During the self-supervised pre-training phase, different pretext tasks are typically designed to allow the model to more effectively learn the intrinsic features and interrelationships within the samples. By performing these pretext tasks, the model can generate pseudo-labels internally to guide its learning, thus achieving self-supervised learning without the need for labeled data. Classic pretext tasks include image inpainting [21], which uses neural networks to repair missing parts by learning texture features; rotation prediction [22], which allows neural networks to grasp the overall features of an image; and jigsaw puzzles [23], where the neural network needs to learn the relative positional features among different pieces for image stitching. These pretext tasks have achieved

good results in instance-level image classification tasks. However, their effectiveness is not ideal for semantic segmentation tasks due to a lack of learning about local features.

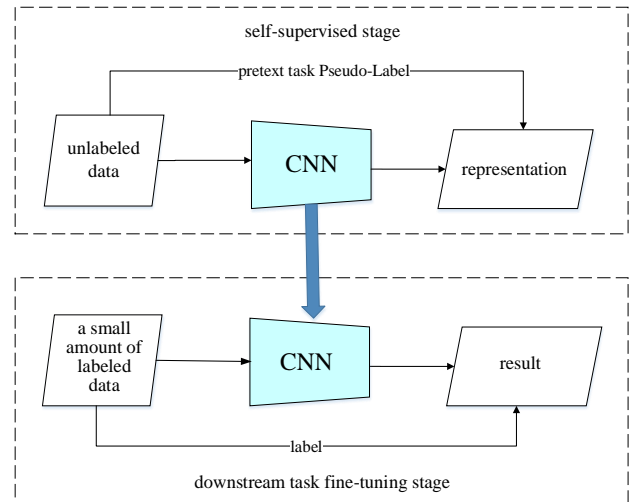


Fig. 1. Schematic diagram of self-supervised model.

C. Contrastive Learning

Contrastive self-supervised learning [24], also referred to as contrastive learning, shows more promising results in the field of remote sensing compared to generative self-supervised learning. The central idea of contrastive learning, a common method of self-supervised learning, is to learn high-level semantic features by contrasting two semantically similar inputs. Specifically, samples are divided into positive and negative pairs, with the aim of drawing positive samples closer while pushing negative samples farther apart, as shown in formula (1):

$$\text{sim}(f(x), f(x^+)) \gg \text{sim}(f(x), f(x^-)) \quad (1)$$

Here, x^+ represents a sample semantically similar to x , thus forming a positive pair with x ; x^- is a sample that is different from x , thereby forming a negative pair with x . sim represents the similarity measure between two pairs of features generated by encoding function f .

Classic examples of contrastive learning include Momentum contrast (MoCo) [25] and SimCLR [26]. MoCo introduces momentum contrast for unsupervised visual representation learning, constructing a dynamic dictionary with a queue and a moving average encoder to improve the effects of contrastive learning. SimCLR presents a simple framework where two different data augmentations of the same image x are generated as a positive pair (x_i and x_j), while the augmented image from a different image y serves as a negative sample. A projection head is added after the encoder to achieve significant results. While both of the above-mentioned models have achieved significant accomplishments in self-supervised learning research, they also exhibit notable limitations. This is because both models utilize pairs of images as positive samples, allowing them to effectively learn overall features of images. However, they lack the ability to learn local features.

Recent years have seen extensive research on image processing based on contrastive learning, such as a method proposed by Krishna et al. for medical image semantic segmentation based on global and local features [27]. Research on self-supervised learning for remote sensing images mainly focuses on instance-level remote sensing scene classification [28][29], given the comparatively limited exploration of pixel-level semantic segmentation in the context of remote sensing images, a triple-branch network architecture is introduced. This architecture facilitates the acquisition of both global and local image features, consequently leading to enhanced semantic segmentation outcomes for remote sensing images.

D. TraIn Method

Deep learning requires a large amount of data support, and the quality of data often has a significant impact on model training. An important measure of data quality is influence, but due to the complexity of models and the growing influence of scale features and datasets, it is challenging to quantify influence. The TraIn method captures changes in predictions when accessing individual training examples by tracking the training process and determines the influence of training examples by assigning influence scores to each.

E. Order Optimization Theory

Order Optimization (OO) is an effective strategy widely used in the industry to solve optimization problems, with its specific solution process shown in Fig. 2.

For a given optimization problem, suppose the set of the "truly best" g solutions is G . However, due to computational resource constraints, the set G cannot be solved from the solution space. Using the order optimization idea, a rough model with simple computations is used to select some solutions from G . All solutions are ranked according to some performance evaluation method provided by the rough model, and the best s solutions are chosen to form the solution set S . In the process of using the rough model, we generally only care about how many of the intersecting parts of sets G and S ($G \cap S$) are genuinely good solutions. The order optimization quantifies the probability that the set S obtained based on the rough model corresponds to $|G \cap S| \geq k$, i.e., the alignment probability (AP). In practice, the alignment probability of sets S and G is often much larger than expected, and the amount of data in set S is often several orders of magnitude smaller than the real solution space, so the order optimization method can typically save at least one order of magnitude of performance evaluation times.

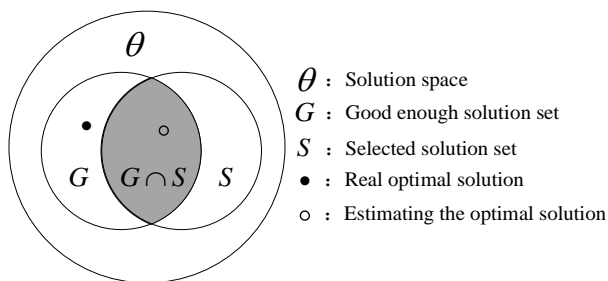


Fig. 2. Schematic diagram of solving sequential optimization theory.

III. PROPOSED METHOD

A. Network Architecture Design for Semantic Segmentation of Remote Sensing Images Based on Self-supervised Learning

With the objective of improving the outcomes of semantic segmentation for remote sensing images using a limited quantity of annotated data, as well as intensifying the acquisition of local small-object characteristics, a triple-branch network architecture known as TBSNet is introduced. As shown in Fig. 3, this network structure includes an image transformation prediction branch, a global contrastive learning branch, and a local contrastive learning branch. The image transformation branch are used to learn the overall features of the image, while the local contrastive learning branch can learn the local features of the image. Each branch performs self-supervised learning in different ways, and then the losses of each branch are summed up as the total loss for adjusting the network parameters.

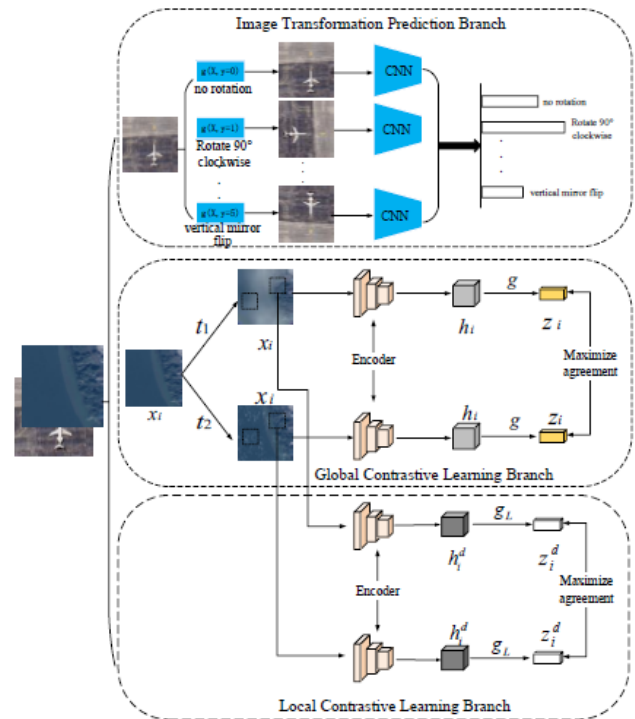


Fig. 3. Schematic diagram of triple-branch network TBSNet.

1) Design of Image Transformation Prediction Branch Learning Strategy: To realize the learning of overall semantic features of images without labels, this branch randomly rotates (e.g., 90°, 180°, 270°) or mirror flips the original image, and feeds the original image and the rotated image into the neural network for transformation type identification. Since remote sensing images have rotation invariance, rotation can help the neural network better understand the concepts described in remote sensing images. Specifically, the aforementioned rotations are defined as a set of discrete geometric transformations $G = \{y_0, y_1, \dots, y_m\}$. One is randomly selected from G and applied to the input image x to get x' , which is

then fed into the network and trained to identify the type of rotation, transforming the image transformation prediction branch into a classification problem. The loss function of the transformation prediction branch can be defined as shown in equation (2):

$$L_{\alpha} = -\sum_{m=1}^M \hat{A}_{(m)} \log P_{(m)} \quad (2)$$

Here, $\hat{A}_{(m)} = \{0,1\}$ represents the one-hot encoding of the basic true value class, and P represents the probability of M different types of geometric transformations. In this branch, geometric transformations are divided into six types, which are clockwise rotation of 90° , 180° , 270° , horizontal left-right mirror flipping, vertical top-bottom mirror flipping, and no rotation.

2) Global Contrastive Learning Branch Learning Strategy Design

a) Remote Sensing Data Enhancement Method Based on Weather Conditions: For the same location, remote sensing images under different weather conditions exhibit variations, yet their deep features remain consistent. Consequently, in the global contrastive learning segment, traditional data augmentation approaches for forming positive sample pairs are eschewed. Instead, the data augmentation method is adjusted to mimic diverse weather conditions at the identical location, aligning with the distinct traits of remote sensing images. Any x_i in $\{x_1, x_2, \dots, x_N\}$ undergoes two different random data augmentations (including simulating clouds, simulating snowflakes, simulating haze, and no augmentation). To simulate cloud layers, this paper adds Perlin noise to the original image and then blurs the cloud layer using a Gaussian filter. To simulate snowflakes, random noise is added to the original image, and then a median filter is used to simulate the snowflake effect. To simulate haze, we utilize a method of overlaying a generated haze layer onto the original image. The haze layer is represented by an array of the same size as the original image. To ensure uniform effect across all color channels, this array is expanded to a three-channel array, with each channel having the same values as the original random array. Each element of the haze layer is multiplied by the haze intensity, and the result is multiplied with each pixel of the original image. This effectively reduces the contrast of the original image in the haze areas. Then, the haze layer is multiplied by the atmospheric brightness and added to the original image, simulating the haze effect. In this paper, the haze intensity is randomly selected between 0.3 and 0.8, and the atmospheric brightness is randomly chosen between 250 and 270.

b) Model design integrating channel and spatial attention mechanisms: The two samples \tilde{x}_i and \hat{x}_i obtained after enhancement are positive samples for each other, and other samples in the same batch are all negative samples. The two positive samples obtained are passed through the encoder-based backbone network f to get the feature vectors \tilde{h}_i and \hat{h}_i , which are then mapped to the contrast loss space through an improved MLP projection head $g(\cdot)$ to get \tilde{z}_i and \hat{z}_i . As shown in Fig. 4, this work introduce the combination of channel

attention mechanism and spatial attention mechanism on the basis of the original projection head, improving the discriminability and expressive power of features. It can also adaptively select important features, which helps to improve the model's generalization ability on different types of remote sensing images. In order to better integrate the channel attention module and the spatial attention module, a convolution layer and ReLU activation function are added. This additional convolution layer can help to further extract features before applying the attention mechanism.

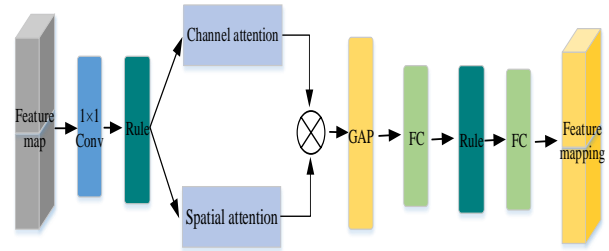


Fig. 4. Schematic diagram of global contrastive learning branch projection head structure.

Lastly, the contrast loss is used to bring positive samples closer, thus learning the geographical features in the image. The contrast loss is defined as follows in equation (3):

$$L_{\beta} = \frac{1}{2N} \sum_{k=1}^N \left(l_{\beta}(\tilde{x}_i, \hat{x}_i) + l_{\beta}(\hat{x}_i, \tilde{x}_i) \right) \quad (3)$$

Where N represents the number of samples in the same batch, l_{β} uses the NT-Xent contrast loss function in SimCLR as shown in equation (4):

$$l_{\beta}(\tilde{x}_i, \hat{x}_i) = -\log \frac{\exp\left(\frac{\text{sim}(\tilde{z}_i, \hat{z}_i)}{\tau}\right)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp\left(\frac{\text{sim}(\tilde{z}_i, z_k)}{\tau}\right)} \quad (4)$$

Here, $\mathbf{1}_{[k \neq i]}$ is an indicator function that equals 1 when $k \neq i$, $\text{sim}(u, v) = \frac{u^T v}{\|u\| \|v\|}$, i.e., it calculates the cosine similarity between u and v . z_k represents the feature vector obtained after the negative sample goes through the projection head, that is, $z_k = g(f(t(x_k)))$. τ is the temperature parameter, which is set to 0.1 in this paper.

3) Local Contrastive Learning Branch Learning Strategy Design: The above two branches can effectively learn the global information of images. However, for semantic segmentation, learning only global features is not enough. A single remote sensing image may contain various objects, and the learning of global features cannot effectively handle small targets. Therefore, the local contrastive learning branch can learn more local information, which is crucial for improving the performance of semantic segmentation. This branch shares the sample after data augmentation with the global contrastive learning branch. It forms positive sample pairs by selecting two local blocks of the same size from the same location in two augmentation images, and takes the local areas of other images in the same batch as negative samples. In this paper, a random selection of a central pixel point and outward expansion method is employed for selecting a local region. To

prevent insufficient edge size, if the selected size is $s \times s$, the central pixel point is chosen within the rectangular region formed by the four points: $\left(\left\lfloor \frac{s}{2} \right\rfloor, \left\lfloor \frac{s}{2} \right\rfloor\right), \left(\left\lfloor \frac{s}{2} \right\rfloor, \left(255 - \left\lfloor \frac{s}{2} \right\rfloor\right)\right), \left(\left(255 - \left\lfloor \frac{s}{2} \right\rfloor\right), \left\lfloor \frac{s}{2} \right\rfloor\right), \left(\left(255 - \left\lfloor \frac{s}{2} \right\rfloor\right), \left(255 - \left\lfloor \frac{s}{2} \right\rfloor\right)\right)$. To avoid excessive duplication of selected regions, pixels are only chosen with odd coordinates. Moreover, to prevent selecting the same region multiple times, each pixel point can only be selected once.

Just like the global contrastive learning branch, for each global contrastive learning image, m local areas are selected for contrast learning. Therefore, for the selected two local areas \widetilde{x}_i^d and \widehat{x}_i^d ($d \in (1, m)$), after going through the encoder, we get the feature vectors \widetilde{h}_i^d and \widehat{h}_i^d , and finally map the local area features to be \widetilde{z}_i^d and \widehat{z}_i^d through the projection head $g_L(\cdot)$ similar to the one in the global contrastive learning branch. As shown in Fig. 5, in this branch, since the sample image has already been cropped to the local area, the attention mechanism is not used in the projection head. Instead, the first fully connected layer in the original projection head is replaced with a convolution layer to retain spatial information and better capture the features within the local area, thus improving the segmentation effect of the model.

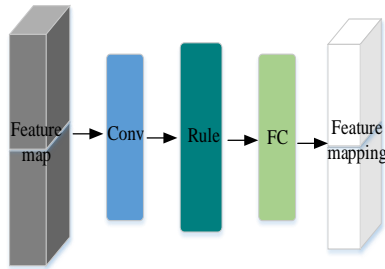


Fig. 5. Schematic diagram of local contrastive learning branch projection head structure.

The contrast loss can be represented as equation (5):

$$L_\gamma = \frac{1}{2N_\gamma} \sum_{i=1}^{N_\gamma} \left(l_c(\widetilde{x}_i^d, \widehat{x}_i^d), l_c(\widetilde{x}_i^d, \widehat{x}_i^d) \right) \quad (5)$$

$$\text{where, } l_c(\widetilde{x}_i^d, \widehat{x}_i^d) = -\log \frac{\exp\left(\frac{\text{sim}(\widetilde{z}_i^d, \widehat{z}_i^d)}{\tau}\right)}{\sum_{k_d \in \Lambda_\gamma^-} \exp\left(\frac{\text{sim}(\widetilde{z}_i^d, z_{k_d}^d)}{\tau}\right)} \quad (6)$$

In this, N_γ represents the number of all local regions in a batch, i.e., $N_\gamma = N \times m$. Λ_γ^- represents the other local regions outside the two local area positive samples.

The total loss of the triple-branch self-supervised network can be represented as shown in equation (7), which is used for the calculation of the TracIn score during optimization.

$$L = L_\alpha + L_\beta + L_\gamma \quad (7)$$

B. Design of Self-supervised Network Architecture Based on Semantic Segmentation of Remote Sensing Images

Since self-supervised pre-training requires a large amount of data as support, but a large amount of data will inevitably increase the calculation amount and consume time cost. Therefore, how to effectively optimize the self-supervised learning algorithm becomes the key to solve the cost problem of self-supervised learning. This paper proposes to optimize the self-supervised learning algorithm by using the TracIn method and sequence optimization theory, achieving the effect of using all data to train the model by only pre-training the model with the top 80% of training points that contribute the most, reducing calculation cost and time cost.

1) *Training point score calculation based on TracIn method:* The TracIn method identifies the overall impact of training examples by tracking the training process. Its principle is as follows: Z represents the sample space, z and z' respectively represent the training point (training sample) and test point (test sample). Given a set of k training points $S = \{z_1, z_2, \dots, z_k \in Z\}$, train the predictor by finding the parameters ω that minimize the training loss $Loss = \sum_{i=1}^k L(\omega, z_i)$ through the iterative optimization process using a training point $z_t \in S$ in iteration t , and update the parameter vector from ω_t to ω_{t+1} . For the training point $z \in S$, the loss reduction caused by the training process for a given test point $z' \in Z$ can be expressed as:

$$TracInIdeal(z, z') = \sum_{t: z_t=z} L(\omega_t, z') - L(\omega_{t+1}, z') \quad (8)$$

By limiting the gradient to a specific gradient descent and substituting the parameter change formula into a first-order approximation and ignoring the high-order term $O(\eta_t^2)$, the following first-order approximation of loss change can be obtained:

$$L(\omega_t, z') - L(\omega_{t+1}, z') \approx \eta_t \nabla L(\omega_t, z') \cdot \nabla L(\omega_t, z) \quad (9)$$

where η_t represents the step length in iteration t .

For a specific training point z , this approximation can approximate the idealized influence by summing this approximation over all iterations where z is used to update the parameters. This first-order approximation is referred to as $TracIn_{TBSNet}$, as shown in equation (10):

$$TracIn_{TBSNet}(z, z') = \sum_{t: z_t=z} \eta_t \nabla L(\omega_t, z') \cdot \nabla L(\omega_t, z) \quad (10)$$

From the above equation, it can be seen that the score of a training point q_i at a test point q_j can be expressed as:

$$tracin_{score_{q_i, q_j}} = TracIn_{TBSNet}(q_i, q_j) \quad (11)$$

In order to verify the difference in contributions among each training point, the scores of each training point on the test point are summed up, and the final score obtained is the total score of this training point, that is:

$$Score_i = \sum_{j=0}^N tracin_{score_{q_i, q_j}} \quad (12)$$

where N represents the number of test points.

2) *Optimization of TBSNet training process based on sequential optimization*: Following the computation of the TraCIn score as outlined earlier, scores can be derived for each training point against the identical set of test points. Based on the idea in sequential optimization that "sequence is more useful than ratio," each training point's TraCIn score can be seen as abstracting each training point into a sortable value. For the triple-branch network model that applies the TraCIn method, the scores of each training point are sorted, and the training points with higher scores are considered to contribute more, while the training points with lower scores are considered to contribute less. Therefore, using only a certain range of higher scoring training points can achieve results comparable to training with all training points. The specific operation process is shown in Fig. 6.

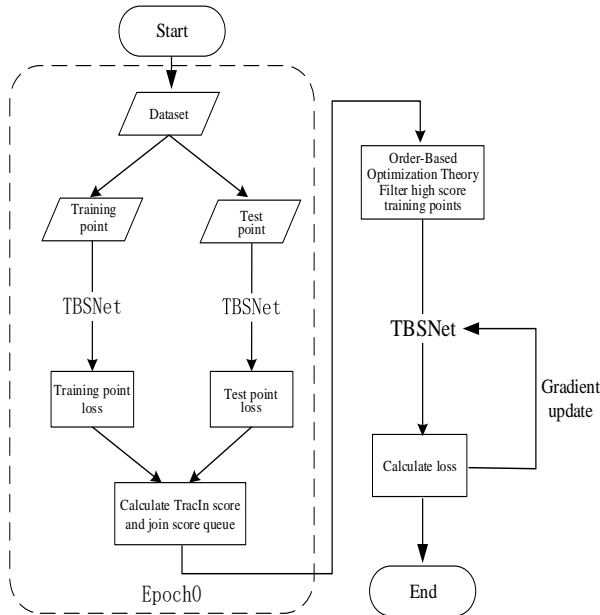


Fig. 6. Schematic diagram of training optimization process.

After the pre-training is completed, the trained neural network is transferred to the downstream task for fine-tuning.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset Selection

The dataset utilized for this study is derived from high-resolution satellite remote sensing images from the southern regions of China. To validate the generalization capability of the proposed method, this dataset amalgamates imagery from various locations and different satellite types. These datasets were uniformly re-annotated into five land cover categories: vegetation, buildings, roads, water bodies, and others. In total, it comprises 12 large-scale RGB original images ranging from 4000×4000 to 8000×8000 pixels.

Due to computational resource constraints, the original remote sensing images were segmented into patches of 256×256 pixels. Additionally, to meet the extensive data

requirements for pre-training, the experimental dataset was augmented using random noise, Gaussian blur, and color transformations, resulting in an enriched dataset. Ultimately, a training sample consisting of 100,000 images was established.

B. Evaluation Metrics

For validating and evaluating the suggested approach in subsequent tasks related to remote sensing image segmentation, the performance metrics employed are Overall Accuracy (OA) and Recall. These metrics are defined in equations (13) and (14) as provided below:

$$OA = \frac{TP}{N} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

Here, TP represents the correctly predicted pixel count, or true positives. FN signifies the incorrectly predicted pixel count, or false negatives. N denotes the total number of pixels.

C. Experimental Setup and Configuration

The experiment was conducted in a Linux environment, with an Intel(R) Xeon(R) Gold 5218R CPU and NVIDIA GeForce RTX 2080Ti 11Gb GPU. Programming was done in Python 3.8 within the PyTorch framework, with Resnet50 as the backbone network and DeepLabV3+ for segmentation. The initial learning rate during the self-supervised stage was set to 0.01, batch size was 32, region size in the local contrast learning branch was 24×24, and six local areas were selected from each image. The pre-training stage was set to run for 500 epochs using the Adam optimizer. The initial learning rate for the fine-tuning stage was set to 0.005, the fine-tuning epoch was set to 50, and the training and validation sets were split in a 7:3 ratio. The selection of training and testing points for calculating the TraCIn score was done per batch, and the ratio of training points to testing points was set at 5:1.

The experiment started with the computation of TraCIn scores for each training point during the first training round. Following the idea of sequential optimization, the top 80% of training points with higher TraCIn scores were selected for further self-supervised pre-training. Finally, after pre-training, the trained model was fine-tuned on a downstream segmentation task using a small amount of labeled data.

D. Comparative Experiment

To validate the effectiveness of our method, we compared it against several mainstream methods, including MoCo v2 [30], which uses a dynamic dictionary for contrastive learning, SimCLR, which uses data augmentation to create positive pairs for contrastive learning, the classic self-supervised learning pretext task of image inpainting, and supervised pre-training models using ImageNet for segmentation. In this paper, we used 0.5%, 1%, 5%, and 10% of the self-supervised pre-training data to fine-tune the downstream task, as shown in Table I. For different models' semantic segmentation experiments on this dataset, some experimental results are shown in Fig. 7.

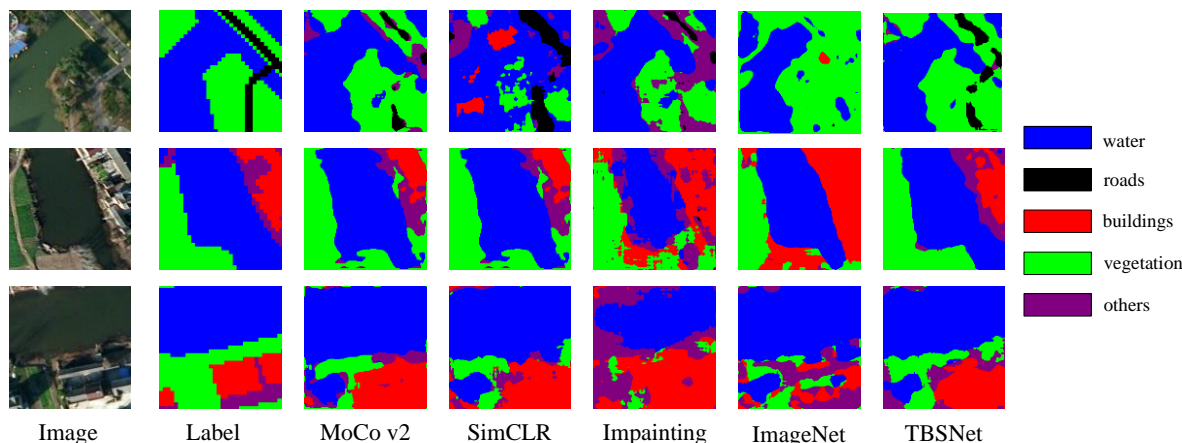


Fig. 7. Comparison experiment effect picture.

TABLE I. COMPARATIVE EXPERIMENTAL RESULTS OF DIFFERENT MODELS

Fine-tune data volume	0.5%		1%		5%		10%	
	OA	Recall	OA	Recall	OA	Recall	OA	Recall
Inpainting	0.2764	0.2593	0.4375	0.4359	0.5158	0.4783	0.5709	0.5256
SimCLR	0.3809	0.3825	0.4604	0.4289	0.5781	0.5452	0.6679	0.6623
MoCo v2	0.3654	0.3577	0.4432	0.3964	0.5295	0.5114	0.6067	0.6008
ImageNet	0.3848	0.3826	0.4128	0.4049	0.5624	0.5551	0.6792	0.6567
Ours(TBSnet)	0.3856	0.3831	0.4721	0.4558	0.5739	0.5584	0.7139	0.7111

Our results show that our method is effective for land cover segmentation in remote sensing images. With only 10% of labeled data used for fine-tuning, the overall accuracy (OA) and recall reached 0.7139 and 0.7111 respectively, representing a significant improvement over advanced self-supervised models such as MoCo v2 and SimCLR.

Analysis of the results reveals that since Inpainting mainly predicts missing areas from the image's context, it often lacks precision for complex remote sensing images, thus performing the worst in the experiments. Both MoCo v2 and SimCLR focus on global features, and their effectiveness is limited due to the lack of learning of local features in remote sensing images. Although ImageNet uses millions of natural images for pre-training, the differences in distribution, texture, and color between remote sensing images and natural images make ImageNet pre-training ineffective for downstream remote sensing image segmentation tasks. Our method performs well in learning both global and local features, demonstrating great application potential worthy of further research and exploration.

E. Ablation Experiments

1) *Ablation experiments on the three-branch network structure:* In the ablation experiments on the triple-branch network structure, 10% of pre-training data was used for fine-tuning. After one round of training, the top 80% of training points based on TracIn scores were selected for training. The following experiments were designed to demonstrate the effectiveness of the proposed method: using only the image rotation prediction branch (Exp1), using only the global

contrast learning branch (Exp2), using only the local contrast learning branch (Exp3), using both the global and local contrast learning branches (Exp4), using the image rotation prediction branch and the global contrast learning branch (Exp5), using the image rotation prediction branch and the local contrast learning branch (Exp6), and using the complete triple-branch network (Exp7). The experiment results are shown in Table II.

TABLE II. EXPERIMENTAL RESULTS OF TRIPLE BRANCH NETWORK ABLATION

	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7
OA	0.5726	0.6593	0.5830	0.6845	0.6507	0.6732	0.7139
Recall	0.5658	0.6494	0.5800	0.6744	0.6457	0.6642	0.7111

The results reveal that a reasonable combination of the three branches is more conducive to the improvement of downstream task performance. The image rotation prediction branch and the global contrast learning branch can learn the overall features of the image. However, due to the lack of local feature learning, they do not achieve the best results, reaching only an overall accuracy of 0.6507 when learning global features only. Without global feature learning, solely learning local features results in an overall accuracy of only 0.5830. The best results are achieved when both global features are learned using the image rotation prediction branch and the global contrast learning branch, and local features are learned using the local contrast learning branch. Compared to the former two scenarios, the overall accuracy is improved by 0.0632 and 0.1309, respectively.

2) *Ablation experiments on training process optimization methods:* This work conducted training using the top 20%, 50%, 80%, 100% of training points based on TracIn scores, and without sequential optimization, randomly selected 80% of training points (Random 80%), to investigate the impact of the TracIn method and sequential optimization on experiment accuracy. The results are shown in Table III. Simultaneously, we explored the relationship between the time consumed and the overall accuracy when pre-training with different data volumes. The results are shown in Fig. 8.

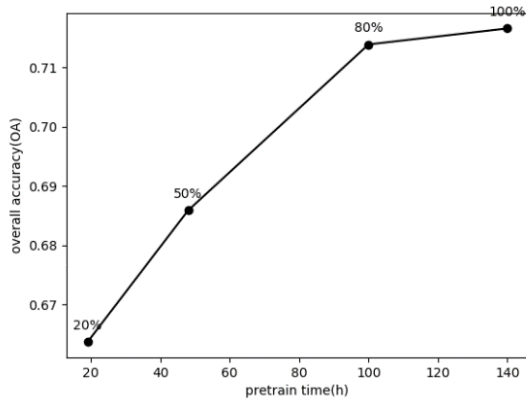


Fig. 8. The impact of different data volumes on time and accuracy.

TABLE III. OPTIMIZATION METHOD ABLATION EXPERIMENTAL RESULTS

	20%	50%	80%	100%	Random 80%
OA	0.6637	0.6859	0.7139	0.7166	0.7048
Recall	0.6599	0.6916	0.7111	0.7183	0.6985

In Fig. 8, with the increase in data volume, both the pre-training time and the overall accuracy increase. However, the ratio of overall accuracy to pre-training time (i.e., the slope of the line) keeps decreasing, indicating that the time required to improve the unit accuracy is increasing. This reflects that those with higher TracIn scores contribute significantly to accuracy improvement. When the data volume reaches the top 80% of TracIn scores, the accuracy is nearly the same as using all pre-training data (i.e., 100%), demonstrating the effectiveness of the optimization method proposed in this paper for reducing data volume. Meanwhile, as shown in Table III, using the top 80% of data based on TracIn scores also improved the results compared to randomly using 80% of the data, verifying the effectiveness of the proposed optimization method.

The experimental results show that applying the training data selected by the TracIn method and sequential optimization to the proposed triple-branch self-supervised network can reduce the data volume by 20% with almost no impact on the experiment accuracy. The reduction in data volume brings about a decrease in time cost. Therefore, for self-supervised learning, the combination of the TracIn method and sequential optimization theory is an optimization format worth considering.

V. CONCLUSION

This study introduces a self-supervised learning-based semantic segmentation technique for remote sensing images. Initially, a triple-branch self-supervised learning network known as TBSNet is developed to capture both global and local features within these images. Subsequently, the TracIn method and sequential optimization theory are employed to enhance the pre-training procedure of the self-supervised learning network, consequently reducing the time and computational resources necessary for pre-training. Ultimately, the pre-trained model is fine-tuned for downstream tasks, culminating in a semantic segmentation model tailored for remote sensing images through self-supervised learning. In comparison to traditional self-supervised models, our experiments reveal varying degrees of enhancement. But there are two notable limitations in this study. First, during the fine-tuning phase, 10% of the labeled data was utilized, thus not achieving a fully unsupervised approach. There remains potential for further reduction in the amount of labeled data used. Secondly, to simultaneously learn global and local features, a triple-branch network collaboration was employed, leading to an increase in the model's size. For future research, under the premise of further reducing labeled data, the goal is to integrate more advanced optimization techniques to develop a lighter self-supervised model and strive to further enhance segmentation accuracy.

REFERENCES

- [1] J. A. Richards, and J. A. Richards, Remote sensing digital image analysis: Springer, 2022.
- [2] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," International journal of multimedia information retrieval, vol. 7, pp. 87-93, 2018.
- [3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation." pp. 3431-3440.
- [4] M. Pastorino, G. Moser, S. B. Serpico, and J. Zerubia, "Semantic segmentation of remote-sensing images through fully convolutional neural networks and hierarchical probabilistic graphical models," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-16, 2022.
- [5] J. M. Alvarez, Y. LeCun, T. Gevers, and A. M. Lopez, "Semantic road segmentation via multi-scale ensembles of learned features." pp. 586-595.
- [6] Zhao, C. Wang, Y. Gao, Z. Shi, and F. Xie, "Semantic segmentation of remote sensing image based on regional self-attention mechanism," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2021.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." pp. 234-241.
- [8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation." pp. 801-818.
- [9] X. Yuan, J. Shi, and L. Gu, "A review of deep learning methods for semantic segmentation of remote sensing imagery," Expert Systems with Applications, vol. 169, pp. 114417, 2021.
- [10] R. Liu, L. Mi, and Z. Chen, "AFNet: Adaptive fusion network for remote sensing image semantic segmentation," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 9, pp. 7871-7886, 2020.
- [11] P. Goyal, D. Mahajan, A. Gupta, and I. Misra, "Scaling and benchmarking self-supervised visual representation learning." pp. 6391-6400.
- [12] W. Li, H. Chen, and Z. Shi, "Semantic segmentation of remote sensing images with self-supervised multitask representation learning," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 6438-6450, 2021.

- [13] S. Albelwi, "Survey on self-supervised learning: auxiliary pretext tasks and contrastive learning methods in imaging," *Entropy*, vol. 24, no. 4, pp. 551, 2022.
- [14] X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks." pp. 6688-6697.
- [15] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks." pp. 286-301.
- [16] Pruthi, F. Liu, S. Kale, and M. Sundararajan, "Estimating training data influence by tracing gradient descent," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19920-19930, 2020.
- [17] Y.-C. Ho, R. S. Sreenivas, and P. Vakili, "Ordinal optimization of DEDS," *Discrete event dynamic systems*, vol. 2, no. 1, pp. 61-88, 1992.
- [18] B. Barlow, "Unsupervised learning," *Neural computation*, vol. 1, no. 3, pp. 295-311, 1989.
- [19] L. Wu, H. Lin, C. Tan, Z. Gao, and S. Z. Li, "Self-supervised learning on graphs: Contrastive, generative, or predictive," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [20] Y. Zheng, M. Jin, Y. Liu, L. Chi, K. T. Phan, and Y.-P. P. Chen, "Generative and contrastive self-supervised learning for graph anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [21] Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting." pp. 2536-2544.
- [22] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," arXiv preprint arXiv:1803.07728, 2018.
- [23] M. Noroozi, and P. Favaro, "Unsupervised learning of visual representations by solving jigsaw puzzles." pp. 69-84.
- [24] P. H. Le-Khac, G. Healy, and A. F. Smeaton, "Contrastive representation learning: A framework and review," *Ieee Access*, vol. 8, pp. 193907-193934, 2020.
- [25] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning." pp. 9729-9738.
- [26] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations." pp. 1597-1607.
- [27] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation with limited annotations," *Advances in neural information processing systems*, vol. 33, pp. 12546-12558, 2020.
- [28] P. Berg, M.-T. Pham, and N. Courty, "Self-supervised learning for scene classification in remote sensing: Current state of the art and perspectives," *Remote Sensing*, vol. 14, no. 16, pp. 3995, 2022.
- [29] Z. Zhao, Z. Luo, J. Li, C. Chen, and Y. Piao, "When self-supervised learning meets scene classification: Remote sensing scene classification based on a multitask learning framework," *Remote Sensing*, vol. 12, no. 20, pp. 3276, 2020.
- [30] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," arXiv preprint arXiv:2003.04297, 2020.

Mechatronics Design and Robotic Simulation of Serial Manipulators to Perform Automation Tasks in the Avocado Industry

Carlos Paredes¹, Ricardo Palomares², Josmell Alva³, José Cornejo⁴

Department of Mechatronic Engineering, Universidad Nacional de Trujillo, Trujillo, Peru^{1, 2, 3}
Universidad Tecnológica del Perú, Lima, Peru⁴

Abstract—Peru is considered one of the principal agroindustrial avocado exporters worldwide. At the beginning of 2022, the volume exported was 8.3% higher than in 2021, so the design and simulation of a pick and place and palletizing cell for agro-exporting companies in the Region of La Libertad was proposed. A methodology was followed that presented a flow diagram of the design of the cell, considering the size of the avocado and the dimensions of the box-type packaging. The forward and inverse kinematics for the Scara T6 and UR10 robots were developed in Matlab according to the Denavit-Hartenberg algorithm, and 3D CAD, dynamic modeling, and trajectory calculation were performed in Solidworks using a "planner" algorithm developed in Matlab, which takes into account the start and end points, maximum speeds, and travel time of each robot. Then, in CoppeliaSim, the working environment of the cell and the robots with their respective configurations are created. Finally, the simulation of trajectories is performed, describing the expected movement, getting the time of the finished task was calculated, where the Scara T6 robot had a working time of 1.18 s and the UR10 of 2.32 s. For 2023 - 2025, its implementation is proposed in the Camposol Company located in the district of Chao - La Libertad, considering the dynamic control of the system.

Keywords—*Mechatronic design; inverse kinematics; dynamic modeling; pick and place; palletizing; Scara robot; universal robot; robot manipulators; path tracking simulation; kinematic control*

I. INTRODUCTION

In 2022, Peru's non-traditional exports grew up by 19.4% compared to 2021, with the agricultural sector accounting for 43.7% of this growth, with avocado as the main product, with a value of 528,727 tons of exported volume [1], [2]. Camposol is considered as the largest Peruvian agro-exporting company, with USD 112,645,000 in shipments [3]. Since 2018 the company developed a strategic expansion plan and purchased 1000 ha in Uruguay [4]. In 2020, Camposol's CEO prioritized the avocado packing process [5]. Various improvement proposals for the packaging process are being studied worldwide. A notable initiative is the development of a 4-degree-of-freedom fruit sorting robot, based on the evaluation of both fruit size and color, using advanced digital image processing techniques. The results have been promising, achieving sorting times of 11.91 seconds for red tomatoes and 11.76 seconds for green ones. Furthermore, it is highlighted that the variation in sorting times is linked to the arrangement of the boxes within the environment. This variability becomes

significant when considering scenarios similar to avocado packing [6]. In another investigation, a control system for a palletizing robot is designed using RobotStudio. This study has demonstrated optimal performance in palletizing tasks using precise input/output (I/O) control logic, which ensures high stability, safety, and efficiency within the simulation environment. This work leads to the proposal to create a virtual environment that simulates real-world conditions for the avocado packing process, elevating it to a TRL5 level [7]. The derivation of the direct and inverse kinematics of the ABB IBR 140 robot was proposed using Denavit-Hartenberg and (DH) analysis and analytical geometric approximations, respectively. The transformation matrices were validated in Matlab and subsequently simulated in RobotStudio to verify their accuracy [8]. A methodology based on the Digital Twin (DT) concept for flexible pick-and-place robotic work cells was also designed to facilitate the development process by providing guidance. This proposal leads to the creation of a design approach for a robotic work cell based on the application of pick-and-place avocado picking and placing for subsequent simulation [9]. A packaging algorithm called Jampack was developed, which has a Failure Recovery Module (FRM) for a robotic manipulator, allowing the system to reach a faster completion of the system [10]. Another algorithm implemented is ResNet-18 for real-time Hass avocado grading, which seeks to achieve non-invasive grading to reduce damage caused by handling. After several processing steps, the image acquisition system achieved an accuracy of 98.72%, a specificity of 98.52%, and a score of 98.08% [11]. However, human-machine collaboration for these applications is still relevant and a topic of ongoing research. This has led to another study aimed at evaluating the avocado harvesting process. After carrying out 41 different tests, a significant increase in yield, measured in the loading zone, from 15% to 80% was found. Also, total harvested production has increased from 23% to 85%, with a minimal increase in human labor load from 1% to only 16% [12]. On the other hand, a soft humanoid hand designed to firmly grip a wide variety of objects, regardless of their morphology, is presented. On this hand, the fingers are constructed with flexible hybrid pneumatic actuators (FHPA). Using a theoretical evaluation model, a balance between the required flexibility and stiffness has been achieved. This innovation offers fast responsiveness and significant gripping force, suitable for fruit picking, product packaging, and handling of fragile objects [13].

This project contributes to Sustainable Development Goals 8 and 9 set by the United Nations, as they focus on achieving higher economic productivity through the utilization of technology and innovation, as well as increasing scientific research and improving technological capacity in industrial sectors, particularly in developing countries, by 2030 [14].

Therefore, the increase in demand for avocado exports poses a challenge if the packaging process is not accelerated, since currently, both at Camposol and other companies, the picking and placing of avocados is performed by workers, resulting in delays in procedures and order deliveries. For this reason, it is proposed to design and simulate a pick and place and palletizing cell, where two robots will perform the process automatically.

The structure of this paper is divided as follows: Section I contains the introduction, which details both the problem and its solution. Section II covers the materials and methods, where the design is based on a methodology that includes kinematic and dynamic analysis, trajectory tracking calculations, and real-time simulation. Section III presents the results derived from the simulation of the robotic cell. Finally, Section IV consolidates the conclusions and future work of the project.

II. MATERIALS AND METHODS

The proposed methodology for designing and simulating the avocado pick and place and palletizing cell, using Matlab and CoppeliaSim, is shown in Fig. 1.

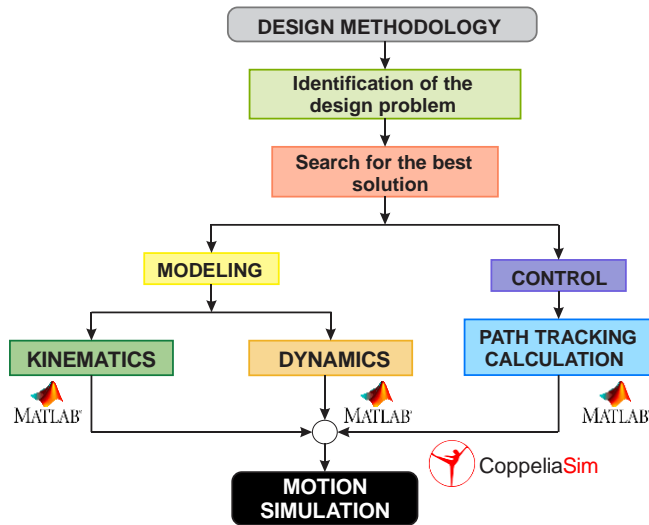


Fig. 1. Flow chart for the design of an avocado pick and place and palletizing cell.

The cell design consists of two tasks that are performed sequentially. The first one is performed by an Epson Scara T6 robot, which picks and places the Hass avocado in a cardboard box (pick and place), using a vacuum suction cup as the final effector, considering the caliber (weight) of the Hass avocado, which varies according to the country to be exported [15], as shown in Table I.

On the other hand, there is a Universal Robots UR10 robot, which picks up the boxes and groups them in rows and columns (palletizing), using a four vacuum suction cup holder

as an end effector, considering the shape, weight, and distribution of the boxes [16], as shown in Table II.

TABLE I. HASS AVOCADO EXPORT CALIBER

Avocado	Avocado Caliber Requirements by Country				Fruit Weight (g)
	USA	Japan	Canada	European Union	
Hass	32	–	12	12	300 – to more
	36	18	14	14	300 – 330
	40	20	16	16	265 – 300
	48	24	18	18	205 – 265
	60	30	20	20	170 – 205
	70	–	22	22	150 – 170
	84	–	24	24	120 – 150

TABLE II. BOX FEATURES

Type of Box	Recommended Avocado Packaging Features		
	Dimensions (mm)	Average Weigh (kg)	Pallet
Cardboard	440 x 338 x 186	11	88 - 96
	406 x 254 x 97	4	228 - 264
Plastic	300 x 500 x 150	10	120

A. Forward and Inverse Kinematic Model

The Scara T6 and UR10 robots are 3 DOF (two rotational and one linear) and 6 DOF (rotational only), respectively [17], [18]. Moreover, their Cartesian reference systems and the motion of each joint are shown in Fig. 2 and Fig. 3 [19].

Eq. (1) presents the T_n^0 homogeneous transformation matrix, which expresses the position and orientation of each robot [20] and [21]. Eq. (2) and (3) represent the basic T_3^0 and T_6^0 transformations, where n is the number of joints.

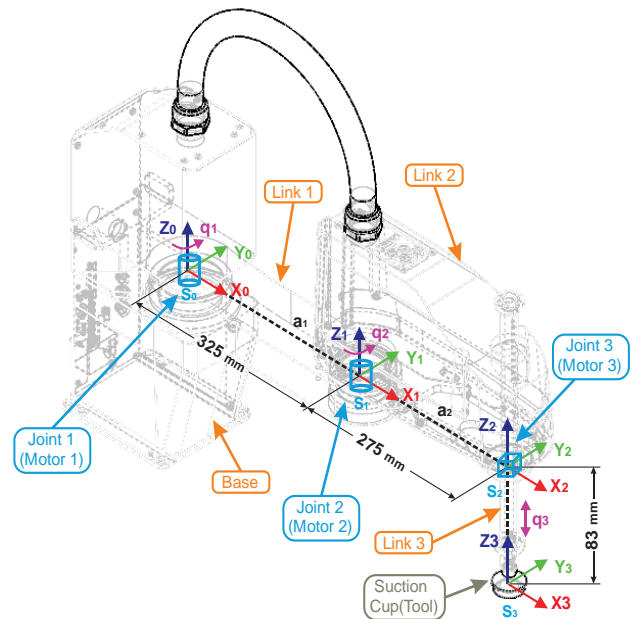


Fig. 2. Cartesian reference system and lengths for the Scara T6 robot.

$${}^0T_n^0 = \begin{bmatrix} n_{nx1} & o_{nx1} & a_{nx1} & P_{nx1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T_3^0 = A_1^0 A_2^1 A_3^2 \quad (2)$$

$$T_6^0 = A_1^0 A_2^1 A_3^2 A_4^3 A_5^4 A_6^5 \quad (3)$$

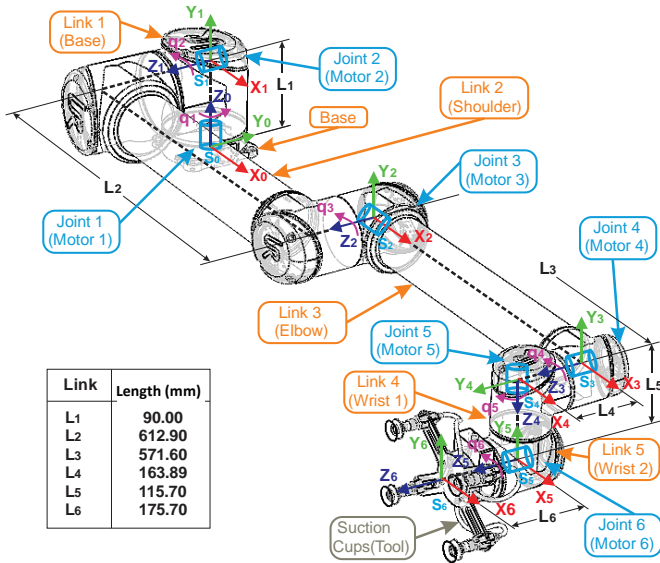


Fig. 3. Cartesian reference system and lengths for the UR10 robot.

The Denavit - Hartenberg (D-H) parameters of the Scara T6 and UR10 robots are shown in Table III and Table IV, respectively [22-24].

TABLE III. D-H PARAMETERS FOR SCARA T6 ROBOT

Joint	Scara T6 Robot			
	θ	d	a	α
1	θ_1	0	a_1	0
2	θ_2	0	a_2	0
3	0	$-\theta_3 - 0.083$	0	0

TABLE IV. D-H PARAMETERS FOR UR10 ROBOT

Joint	UR10 Robot			
	θ	d	a	α
1	q_1	L_1	0	$\pi/2$
2	q_2	0	L_2	0
3	q_3	0	L_3	0
4	q_4	L_4	0	$\pi/2$
5	q_5	L_5	0	$-\pi/2$
6	q_6	L_6	0	0

For the Scara T6 robot by multiplying all the transformation matrices in equation (2), the resulting matrix is:

$$T_3^0 = \begin{bmatrix} n_{3x} & o_{3x} & a_{3x} & P_{3x} \\ n_{3y} & o_{3y} & a_{3y} & P_{3y} \\ n_{3z} & o_{3z} & a_{3z} & P_{3z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Where the position of the end-effector is described by the position vector P_{3x1} (P_{3x}, P_{3y}, P_{3z}), then:

$$P_{3x} = a_2 C_{12} + a_1 C_1 \quad (5)$$

$$P_{3y} = a_2 S_{12} + a_1 S_1 \quad (6)$$

$$P_{3z} = -\theta_3 - 0.083 \quad (7)$$

The inverse kinematics for the Scara T6 robot was solved using the position vector P_{3x1} , obtaining:

$$\theta_1 = 2\text{atan}(t_1) \quad (8)$$

$$\theta_2 = \text{atan}\left(\pm\sqrt{1-t_2^2}/t_2\right) \quad (9)$$

$$\theta_3 = -P_{3z} \quad (10)$$

$$t_1 = \frac{-2P_{3y}a_1 \pm \sqrt{(2P_{3x}a_1)^2 + (2P_{3y}a_1)^2 - (a_2^2 - a_1^2 - P_{3x}^2 - P_{3y}^2)^2}}{a_2^2 - a_1^2 - P_{3x}^2 - P_{3y}^2 - 2P_{3x}a_1} \quad (11)$$

$$t_2 = \frac{P_{3x}^2 + P_{3y}^2 - a_1^2 - a_2^2}{2a_1a_2} \quad (12)$$

On the other hand, for Robot UR10 when multiplying the transformation matrices of equation (3), the resulting matrix is:

$$T_6^0 = \begin{bmatrix} n_{6x} & o_{6x} & a_{6x} & P_{6x} \\ n_{6y} & o_{6y} & a_{6y} & P_{6y} \\ n_{6z} & o_{6z} & a_{6z} & P_{6z} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

Where the position of the end effector is described by the vector (P_{6x}, P_{6y}, P_{6z}) , then:

$$P_{6x} = L_6(S_1 C_5 - C_1 C_{234} S_5) + L_4 S_1 + L_2 C_1 C_2 + L_5 C_1 S_{234} + L_3 C_1 (C_2 C_3 - S_2 S_3) \quad (14)$$

$$P_{6y} = L_2 S_1 C_2 - L_4 C_1 - L_6 (C_1 C_5 - S_1 C_{234} S_5) + L_5 S_1 S_{234} + L_3 S_1 (C_2 C_3 - S_2 S_3) \quad (15)$$

$$P_{6z} = L_1 + L_3 S_{23} + L_2 S_2 - L_5 C_{23} C_4 - S_{23} S_4 - L_6 S_5 (C_{23} S_4 + S_{23} C_4) \quad (16)$$

The inverse kinematics for the UR10 robot was solved using the iterative Gauss-Newton algorithm, which seeks to find the parameter values through an iterative multiplication of matrices [25, 26]. From the transformation matrix of Eq. (3), it follows that:

$$A_6^0 = T_6^0 \quad (17)$$

$$A_6^0 = A_1^0 A_2^1 A_3^2 A_4^3 A_5^4 A_6^5 \quad (18)$$

$$T_6^1(q_1, q_5, q_6) = (A_1^0)^{-1} T_6^0 \quad (19)$$

$$A_4^0 = A_1^0 A_2^1 A_3^2 A_4^3 \quad (20)$$

$$T_4^0(q_2, q_3, q_4) = T_6^0 (A_6^5)^{-1} (A_4^3)^{-1} \quad (21)$$

Then, the parameters are shown below:

$$q_1 = \text{atan2}(L_4, \text{sqrt}((P_{6x} - L_6 a_{6x})^2 + (L_6 a_{6y} - P_{6y})^2 - L_4^2)) - \text{atan2}(L_6 a_{6y} - P_{6y}, P_{6x} - L_6 a_{6x}) \quad (22)$$

$$q_2 = \text{atan2}(r_2, r_1) - \text{atan2}(L_3 S_3, L_2 + L_3 C_3) \quad (23)$$

$$q_3 = \text{atan2} \left(\sqrt{1 - \left(\frac{r_1^2 + r_2^2 - L_2^2 - L_3^2}{2L_2L_3} \right)^2}, \frac{r_1^2 + r_2^2 - L_2^2 - L_3^2}{2L_2L_3} \right) \quad (24)$$

$$q_4 = q_{234} - q_3 - q_2 \quad (25)$$

$$q_5 = \text{atan2} \left(\text{sqrt} \left((n_{6x}S_1 - n_{6y}C_1)^2 + (o_{6x}S_1 - o_{6y}C_1)^2 \right), a_{6x}S_1 - a_{6y}C_1 \right) \quad (26)$$

$$q_6 = \text{atan2} \left(\frac{-o_{6x}S_1 + o_{6y}C_1}{S_5}, \frac{n_{6x}S_1 - n_{6y}C_1}{S_5} \right) \quad (27)$$

$$q_{234} = \text{atan2}(n_{6z}C_5C_6 - a_{6z}S_5 - o_{6z}C_5S_6, o_{6z}C_6 + n_{6z}S_6) \quad (28)$$

$$r_1 = \text{sqrt} \left(\left(\text{sqrt} \left((P_{6x} - L_6a_{6x} + L_5o_{6x}C_6 + L_5n_{6x}S_6)^2 + (P_{6y} - L_6a_{6y} + L_5o_{6y}C_6 + L_5n_{6y}S_6)^2 \right) - L_4^2 \right) \right) \quad (29)$$

$$r_2 = P_{6z} - L_6a_{6z} + L_5o_{6z}C_6 + L_5n_{6z}S_6 - L_1 \quad (30)$$

B. Dynamic Model

The dynamic model for the Scara T6 and UR10 robot was developed using the Newton - Euler formulation, analyzing the geometric relationships between each link, the position vector concerning the i -nth system (ri0pi), and the center of the mass vector of element i concerning its system (ri0si) [27,28].

Eq. (31) and (32) allowed obtaining forces and moments for each element $i = 1, 2, 3, \dots, n$.

$$f_i = F_i + f_{i+1} = m_i \bar{a}_i + f_{i+1} \quad (31)$$

$$n_i = n_{i+1} + \dot{p}_i x f_{i+1} + (\dot{p}_i + \bar{s}_i) x F_i + N_i \quad (32)$$

The physical parameters of each link such as length, mass, center of mass, and inertia tensor, were calculated using CAD modeling designed in SolidWorks, as shown in Fig. 4 and Fig.5 [29-31]. The following steps were followed to obtain it:

- 3D modeling of each element of the Scara T6 and UR10 robots using SolidWorks.
- Generate a separate solid model for each robot link, which should incorporate all significant features, including the motor, bearings, and any other components that contribute to the weight of the link.
- Assign the corresponding material to each component of the robots.
- Identify the primary coordinate system of each link, ensuring alignment with the Denavit - Hartenberg coordinate system. In the event of non-alignment, a new coordinate system should be introduced. Subsequently, upon accessing the properties section in SolidWorks, a summary of the physical property values is displayed, such as mass, center of mass (riosi), and moment of inertia of the link relative to the assigned output coordinate system.
- Finally, use the measurement tool in SolidWorks to draw lines and obtain the length value and the position

vector relative to the i -th coordinate system (riopi) of each link.

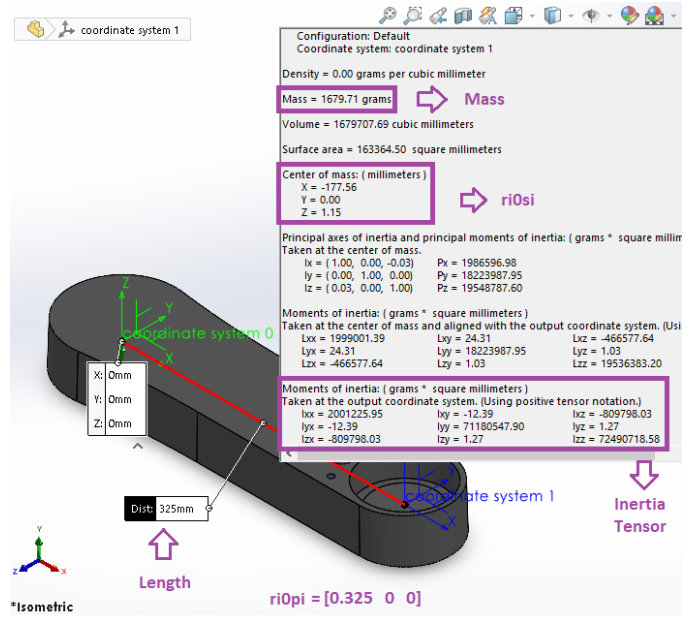


Fig. 4. Physical properties of scaraT6 robot link 1.

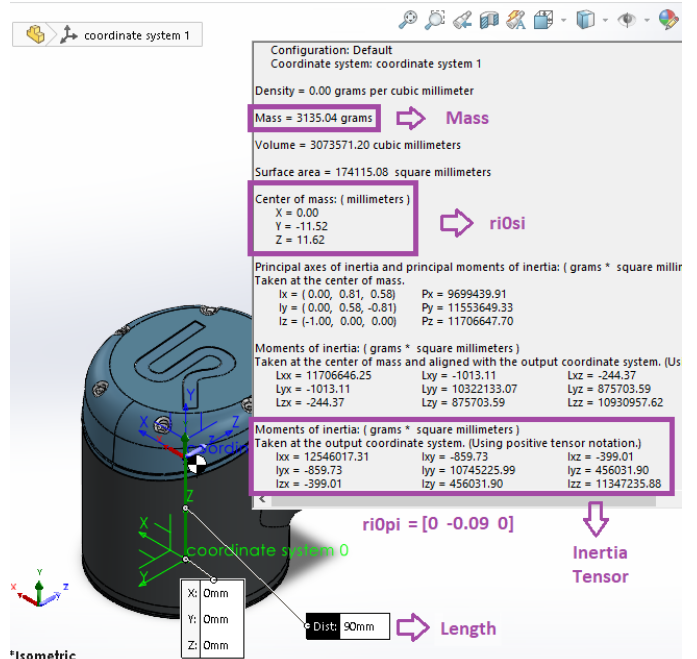


Fig. 5. Physical properties of UR10 robot link 1.

Then the torque for each joint is expressed in equation (33):

$$\tau_i = \begin{cases} n_i^T Z_{i-1} + b_i \dot{q}_i \\ f_i^T Z_{i-1} + b_i \dot{q}_i \end{cases} \quad (33)$$

Where b_i is the coefficient of viscous friction of each joint [32,33].

The parameters obtained from the links were used for the dynamic modeling of the Scara T6 and UR10 robots [34 and [35] which are shown in Table V and Table VI.

TABLE V. PARAMETERS FOR THE DYNAMICS OF THE SCARA T6 ROBOT

n	Scara T6 Robot				
	Length (mm)	Mass (kg)	Inertia Tensor (kgm ²)	riopi (m)	riosi (m)
1	325	1.68	$\begin{bmatrix} 0 & 0 & -0 \\ 0 & 0.07 & 0 \\ -0 & 0 & 0.07 \end{bmatrix}$	$\begin{bmatrix} 0.325 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.18 \\ 0 \\ 0.001 \end{bmatrix}$
2	275	7.54	$\begin{bmatrix} 0.15 & 0 & -0.15 \\ 0 & 0.4 & 0 \\ -0.15 & 0 & 0.265 \end{bmatrix}$	$\begin{bmatrix} 0.275 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.16 \\ 0 \\ 0.126 \end{bmatrix}$
3	-	0.08	$\begin{bmatrix} 0.006 & 0 & 0 \\ 0 & 0.006 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ -0.083 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0.246 \end{bmatrix}$

TABLE VI. PARAMETERS FOR THE DYNAMICS OF THE UR10 ROBOT

n	UR10 Robot				
	Length (mm)	Mass (kg)	Inertia Tensor (kgm ²)	riopi (m)	riosi (m)
1	90	3.14	$\begin{bmatrix} 0.013 & 0 & 0 \\ 0 & 0.01 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.09 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -0.0115 \\ 0.0116 \end{bmatrix}$
2	613	8.89	$\begin{bmatrix} 0.28 & 0 & -0.54 \\ 0 & 1.89 & 0 \\ -0.54 & 0 & 1.62 \end{bmatrix}$	$\begin{bmatrix} 0.613 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.3618 \\ 0 \\ 0.17034 \end{bmatrix}$
3	572	4.75	$\begin{bmatrix} 0.02 & 0 & -0.08 \\ 0 & 0.69 & 0 \\ -0.08 & 0 & 1.63 \end{bmatrix}$	$\begin{bmatrix} 0.572 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} -0.3147 \\ 0 \\ 0.05142 \end{bmatrix}$
4	164	0.74	$\begin{bmatrix} 0.001 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.164 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -0.0076 \\ 0.00974 \end{bmatrix}$
5	116	0.74	$\begin{bmatrix} 0.001 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 0.01 \end{bmatrix}$	$\begin{bmatrix} 0 \\ -0.12 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.0076 \\ 0.00974 \end{bmatrix}$
6	176	0.41	$\begin{bmatrix} 0.003 & 0 & 0 \\ 0 & 0.003 & 0 \\ 0 & 0 & 0.001 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0.176 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.00477 \\ -0.0792 \end{bmatrix}$

From Eq. (31-33) considering b_i negligible and replacing the calculated parameters, the dynamic equations for the Scara T6 robot were obtained:

$$\tau_{1,3} = (1.294 + I_{ext} + 0.181m_{ext} + 0.601C_2 + 0.179m_{ext}C_2)qpp_1 + (0.38 + I_{ext} + 0.076m_{ext} + 0.301C_2 + 0.089m_{ext}C_2)qpp_2 - (0.301S_2 + 0.089m_{ext}S_2)qp_2^2 - (0.061S_2 + 0.179m_{ext}S_2)qp_1qp_2 \quad (34)$$

$$\tau_{2,3} = (0.38 + I_{ext} + 0.076m_{ext} + 0.301C_2 + 0.089m_{ext}C_2)qpp_1 + (0.38 + I_{ext} + 0.076m_{ext})qpp_2 + (0.301S_2 + 0.089m_{ext}S_2)qp_1^2 \quad (35)$$

$$\tau_{3,3} = (m_{ext} + 0.083)(qpp_3 + 9.81) \quad (36)$$

In the same way, the dynamic equations for the UR10 robot were obtained by solving Eq. (31), (32), and (33). By resolving

the dynamics of both robots, the Walker-Orin algorithm (inverse kinematics) is applied to validate the results and potentially facilitate dynamic control [36], [37].

C. Path Tracking Calculation

Kinematic control was performed using Matlab software to develop the desired movements according to the assigned tasks. The developed algorithm "planner" includes the starting and braking times of each motor, as well as the maximum speeds (rad/s) of each robot according to the manufacturer's data sheet [38, 39], for the planning of a smooth trajectory a 4-3-4 interpolator was implemented between the start and end point in the joint coordinates, returning the position, velocity, and acceleration matrices as a result [40-42].

Taking the initial and final points of the path taken by each robot (P_{nx1}), and considering their respective orientations ($R_{n \times n}$), the position, velocity, and acceleration graphs were obtained, as well as their trajectories in the 3D plane, which are shown in Fig. 6 to 9 [43-45].

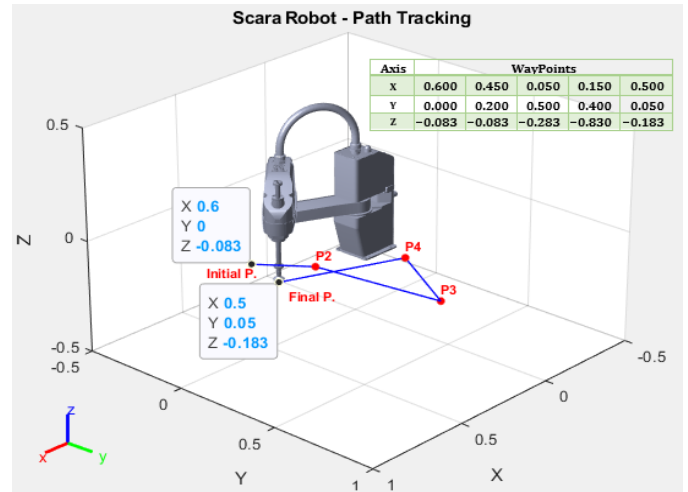


Fig. 6. Path tracking in matlab for scara T6 robot.

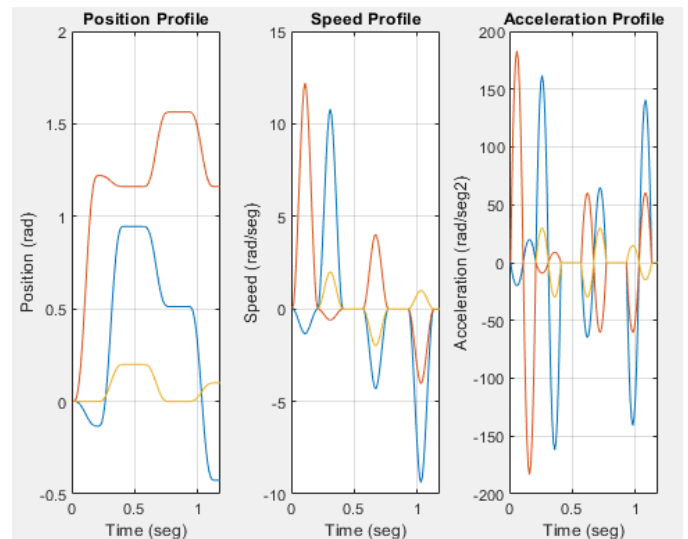


Fig. 7. Trajectory plot in matlab for scara T6 robot.

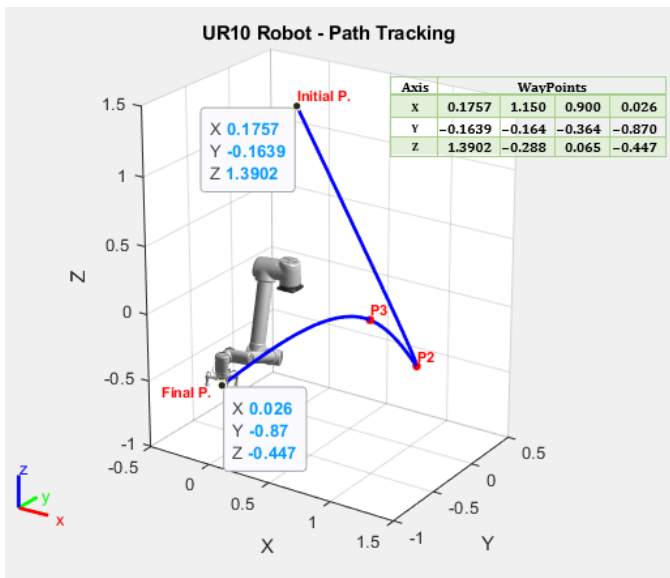


Fig. 8. Path tracking in matlab for UR10 robot.

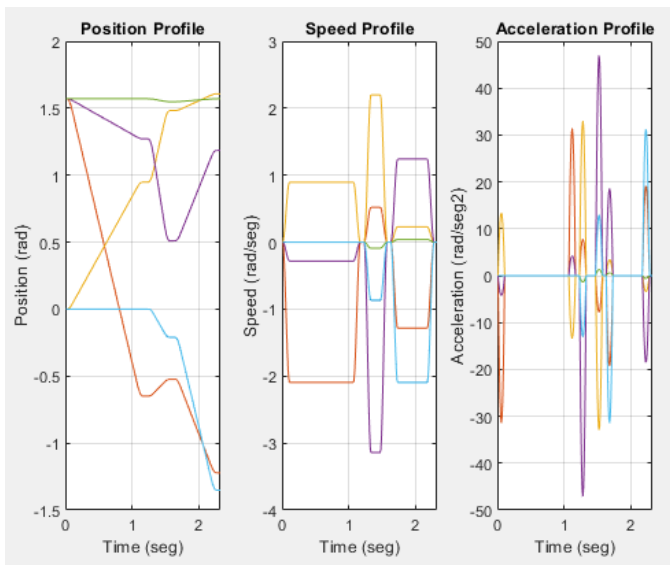


Fig. 9. Trajectory plot in matlab for UR10 robot.

D. Pick and Place and Palletizing Cell Simulation

The data of the joint positions obtained with the kinematic control were saved in matrices called "math_scara" and "math_ur10" respectively, to perform a real simulation in the CoppeliaSim environment [46], [47].

Facilities of the agro-exporting company Camposol were replicated using equipment such as crates, protective enclosures, smooth and roller conveyors provided by CoppeliaSim, along with those designed in SolidWorks, including the SCARA T6 and UR10 robots and the avocado [48], [49]

The designs generated in SolidWorks were then saved in URDF format, including their respective reference systems following the Denavit-Hartenberg method, to enable their visualization in the CoppeliaSim simulation environment to create a highly realistic virtual environment [50-52].

The SolidWorks URDF CAD files of the Scara T6 and UR10 robots were imported into CoppeliaSim, with the relevant parameters for real-time simulation, such as maximum torque (Nm) and maximum speed (°/s), being configured based on the manufacturer's datasheet. This process is illustrated in Fig. 10 [53-55].

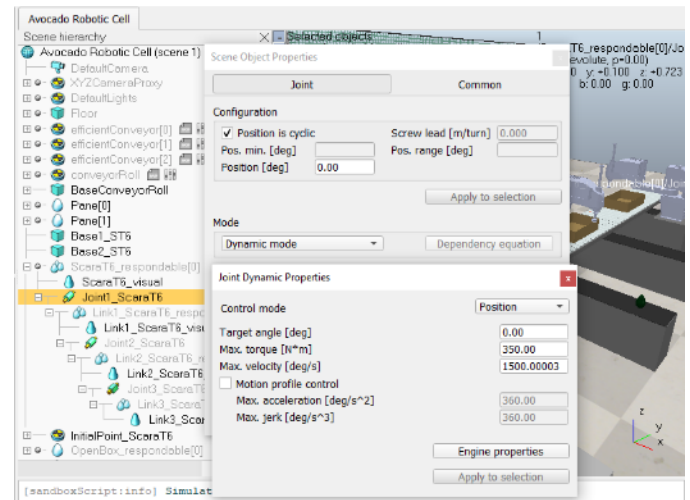


Fig. 10. Creation of the working environment and configuration of the Scara T6 y UR10 robots.

III. TESTS AND RESULTS

Fig.11 shows the Scara T6 robot picking up the avocado, starting from its resting position (a), and then placed in a cardboard box (b), at a maximum speed of 1500 °/s in its rotational joints and of 10 m/s in its prismatic joint, with a cycle time of 0.49 s. The task is performed until 20 Hass avocados of 22 – 24 calibers are placed in each box (c), then the suction cup type tool is activated, taking the box to the pallet to be grouped in three rows and three columns (d), at a maximum speed of 120 °/s for the base and shoulder joints, 180 °/s for the elbow and wrist, with a cycle time of 0.5 s. The task is executed up to palletizing 90 boxes.

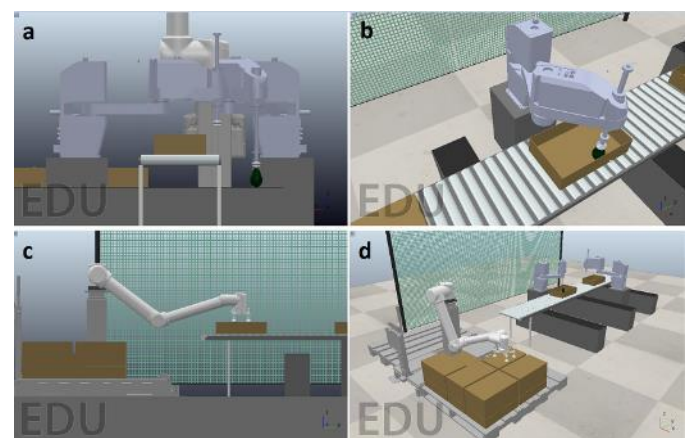


Fig. 11. Simulation in CoppeliaSim for scara T6 and UR10 robots.

The cell movements for pick and place and palletizing were simulated in CoppeliaSim. Table VII shows the working time of each path for both robots. A total working time of 1.18 s was obtained for the Scara T6 robot and 2.32 s for the UR10 robot.

For maximum range position $[\pi/2 \text{ rad/s}, \pi/2 \text{ rad/s}, 0.2 \text{ m}]$ for the Scara T6 robot and $[0, 0, 0, 0, 0, 0, 0]$ rad/s for the UR10, the maximum torques $[360.70, 146.30, 0.376]$ Nm and $[62.88, 116.20, 40.56, 9.25, 6.73, 4.34]$ Nm was obtained for each joint of both robots, as shown in Fig. 12.

TABLE VII. WORKING TIME IN EACH STATION

Path	Working Time	
	Scara T6 Robot (s)	UR10 Robot (s)
1 - 2	0.20	1.22
2 - 3	0.36	0.40
3 - 4	0.36	0.70
4 - 5	0.26	-
TOTAL	1.18	2.32

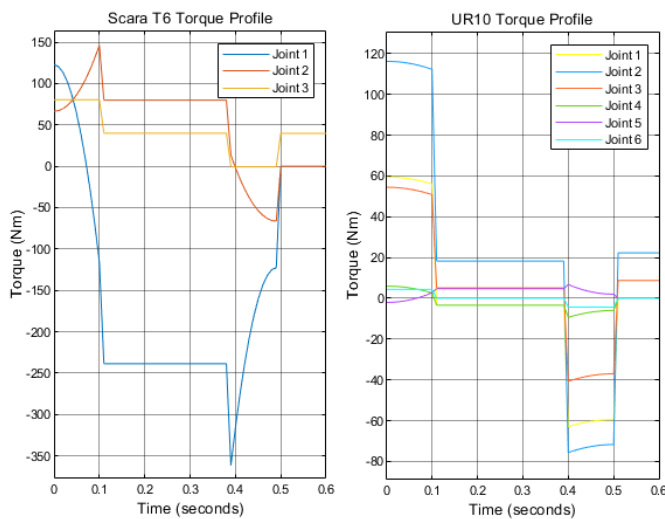


Fig. 12. Torque plot in simulink for scara T6 y UR10 robots.

IV. CONCLUSIONS AND FUTURE WORK

The mathematical models obtained using the Denavit-Hartenberg algorithm of the Scara T6 and UR10 robots were validated with the inverse kinematics tests developed in Matlab, and the models describing the inverse Newton Euler dynamics of the Scara T6 and UR10 robots were validated with the forward dynamics of Walker Orin developed in Matlab. With the "planner" algorithm developed, it was possible to follow the waypoints with smooth movements in each of the joints of the Scara T6 and UR10 robot, as evidenced in the graphs of the positioning, velocity, and acceleration profiles obtained. This significantly influences the accurate execution of pick and place as well as palletizing tasks assigned to the robots within the simulation environment.

The movements and constraints of the global simulation in CoppeliaSim allowed us to get efficient results for the serial work of the Scara T6 robot and the UR10 in the pick and place and palletizing cell with a cycle time of 3.5 s. According to the torque graphs obtained with Simulink, the maximum values for the Scara T6 robot and the UR10 robot showed a minimum difference of 3.06 % and 2.35 %, respectively, concerning the manufacturer's datasheet that was input in the dynamic

configurations of each joint in CoppeliaSim for each robot. This set of actions, in turn, leads to a significant reduction in the time required for the avocado packing process compared to human labor. This is especially relevant since more human personnel would be needed to achieve equivalent performance.

During 2023 – 2025, it is proposed to implement the pick and place tasks in a palletizing cell at the Camposol company located in the district of the Chao, La Libertad region, including a dynamic controller for both robots. In addition, this work not only complements previous research but also opens the door to the possibility of replicating this same application at an industrial level throughout South America and in countries that also produce and export avocados. This allows for the potential to export avocados on a large scale. Thus, meeting increased demand would no longer be an unattainable challenge.

REFERENCES

- [1] ComexPerú, "Las exportaciones de palta cayeron un 9.5% entre enero y agosto de este año," *ComexPerú*, 2022. <https://www.comexperu.org.pe/articulo/las-exportaciones-de-palta-cayeron-un-95-entre-enero-y-agosto-de-este-año>
- [2] R. Gestión, "Las razones detrás de la caída de las exportaciones de palta peruana," *Redacción Gestión*, 2022. <https://gestion.pe/economia/las-razones-detras-de-la-caida-de-las-exportaciones-de-palta-peruana-comex-noticia/>
- [3] Blue Berries, "Ranking de las diez principales empresas agroexportadoras peruanas," *Blue Berries*, 2022. <https://blueberriesconsulting.com/ranking-de-las-diez-principales-empresas-agroexportadoras-peruanas>
- [4] Redagrícola, "Camposol compra 1,000 ha en Uruguay," *Redagrícola*, 2018. <https://www.redagricola.com/pe/camposol-compra-1000-ha-uruguay>
- [5] S. Rosales, "Camposol priorizará packing de palta el 2021 y a partir del 2022 continuará su expansión," *Redacción Gestión*, 2020. <https://gestion.pe/economia/empresas/camposol-el-proximo-año-priorizaremos-packing-de-palta-y-desde-2022-continuaremos-con-expansion-agroexportaciones-noticia/>
- [6] T. Dewi, P. Risma, and Y. Oktarina, "Fruit sorting robot based on color and size for an agricultural product packaging system," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1438–1445, 2020, doi: 10.11591/eei.v9i4.2353.
- [7] X. Chen, Y. Zhou, B. Yang, X. Miao, Y. Li, and M. Zhang, "Designing Control System of Palletizing Robot Based on RobotStudio," *Journal of Physics: Conference Series*, vol. 2402, no. 1, 2022, doi: 10.1088/1742-6596/2402/1/012039.
- [8] M. Almaged, "Forward and Inverse Kinematic Analysis and Validation of the ABB IRB 140 Industrial Robot," *International Journal of Electronics, Mechanical and Mechatronics Engineering*, vol. 7, no. 2, pp. 1383–1401, 2017, doi: 10.17932/iau.ijemme.21460604.2017.7/2.1383-1401.
- [9] B. Tipary and G. Erdős, "Generic development methodology for flexible robotic pick-and-place workcells based on Digital Twin," *Robotics and Computer-Integrated Manufacturing*, vol. 71, 2021, doi: 10.1016/j.rcim.2021.102140.
- [10] M. Agarwal, S. Biswas, C. Sarkar, S. Paul, and H. S. Paul, "Jampacker: An Efficient and Reliable Robotic Bin Packing System for Cuboid Objects," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 319–326, 2021, doi: 10.1109/LRA.2020.3043168.
- [11] O. J. V. Ramirez, J. E. Cruz De La Cruz, W. A. M. MacHaca, "Agroindustrial Plant for the Classification of Hass Avocados in Real-Time with ResNet-18 Architecture," *2021 5th International Conference on Robotics and Automation Sciences, ICRAS 2021*, pp. 206–210, 2021, doi: 10.1109/ICRAS52289.2021.9476659.
- [12] J. P. Vásquez, F. A. Auat Cheein, "Workload and production assessment in the avocado harvesting process using human-robot

- collaborative strategies,” *Biosystems Engineering*, vol. 223, pp. 56–77, 2022, doi: 10.1016/j.biosystemseng.2022.08.010.
- [13] X. Liu, Y. Zhao, D. Geng, S. Chen, X. Tan, C. Cao, “Soft Humanoid Hands with Large Grasping Force Enabled by Flexible Hybrid Pneumatic Actuators,” *Soft Robotics*, vol. 8, no. 2, pp. 175–185, 2021, doi: 10.1089/soro.2020.0001.
- [14] UNITED NATIONS, “Sustainable Development Goals,” *UNITED NATIONS*, 2023. <https://www.un.org/sustainabledevelopment/>
- [15] Aguacate Hass y Méndez, “Tabla de calibres aguacate de exportación,” *simpaq*, 2022. <https://www.simpaq.mx/producto>
- [16] Unifrutti, “Embalaje / paletizaje,” *unifrutti*, 2022. <https://www.unifrutti.com/productos/paltas/embalaje-paletizaje/>
- [17] J. Huanca, J. Zamora, J. Cornejo, and R. Palomares, “Mechatronic Design and Kinematic Analysis of 8 DOF Serial Robot Manipulator to Perform Electrostatic Spray Painting Process on Electrical Panels,” *Proceedings of the 2022 IEEE Engineering International Research Conference, EIRCON 2022*, 2022, doi: 10.1109/EIRCON56026.2022.9934104.
- [18] J. Cornejo, J. Palacios, A. Escobar, and Y. Torres, “Mechatronics Design and Kinematic Simulation of UTP-ISR01 Robot with 6-DOF Anthropomorphic Configuration for Flexible Wall Painting,” *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*, 2022, doi: 10.1109/ICEEICT53079.2022.9768599.
- [19] D. He, F. Liu, and F. Wang, “Optimal Design of Industrial Robot Kinematics Algorithm,” *Journal of Physics: Conference Series*, vol. 1624, no. 4, pp. 3–7, 2020, doi: 10.1088/1742-6596/1624/4/042029.
- [20] S. Kana, J. Gurnani, V. Ramanathan, S. H. Turlapati, M. Z. Ariffin, and D. Campolo, “Fast Kinematic Re-Calibration for Industrial Robot Arms,” *Sensors*, vol. 22, no. 6, 2022, doi: 10.3390/s22062295.
- [21] G. Gao, G. Sun, J. Na, Y. Guo, and X. Wu, “Structural parameter identification for 6 DOF industrial robots,” *Mechanical Systems and Signal Processing*, vol. 113, pp. 145–155, 2018, doi: 10.1016/j.ymsp.2017.08.011.
- [22] H. Chen, N. Zhou, and R. Wang, “Design and Dimensional Optimization of a Controllable Metamorphic Palletizing Robot,” *IEEE Access*, vol. 8, pp. 123061–123074, 2020, doi: 10.1109/ACCESS.2020.3007707.
- [23] O. J. V. Ramirez, J. E. Cruz De La Cruz, and W. A. M. MacHaca, “Agroindustrial Plant for the Classification of Hass Avocados in Real-Time with ResNet-18 Architecture,” *2021 5th International Conference on Robotics and Automation Sciences, ICRAS 2021*, pp. 206–210, Jun. 2021, doi: 10.1109/ICRAS52289.2021.9476659.
- [24] J. Cornejo, V. Cruz, F. Carrillo, R. Cerda, and E. R. Sanchez Penadillo, “Mechatronics Design and Kinematic Simulation of SCARA Robot to improve Safety and Time Processing of Covid-19 Rapid Test,” *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*, 2022, doi: 10.1109/ICEEICT53079.2022.9768506.
- [25] O. Mejia, D. Nunez, J. Razuri, J. Cornejo, and R. Palomares, “Mechatronics Design and Kinematic Simulation of 5 DOF Serial Robot Manipulator for Soldering THT Electronic Components in Printed Circuit Boards,” *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*, 2022, doi: 10.1109/ICEEICT53079.2022.9768447.
- [26] J. Cornejo, R. Palomares, M. Hernandez, D. Magallanes, and S. Gutierrez, “Mechatronics Design and Kinematic Simulation of a Tripterion Cartesian-Parallel Agricultural Robot Mounted on 4-Wheeled Mobile Platform to Perform Seed Sowing Activity,” *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies, ICEEICT 2022*, 2022, doi: 10.1109/ICEEICT53079.2022.9768422.
- [27] E. Re and A. Todas, “Robot Epson SCARA T6 All-in-One.” pp. 21–23.
- [28] U. Robots, “UR10e Technical Specifications.” p. 177.
- [29] M. A. González-Palacios, M. A. García-Murillo, and M. González-Dávila, “A novel tool to optimize the performance of SCARA robots used in pick and place operations,” *Journal of Mechanical Science and Technology*, vol. 35, no. 10, pp. 4715–4726, Oct. 2021, doi: 10.1007/S12206-021-0937-X.
- [30] H. Chanal *et al.*, “Geometrical defect identification of a SCARA robot from a vector modeling of kinematic joints invariants,” *Mechanism and Machine Theory*, vol. 162, p. 104339, 2021, doi: 10.1016/j.mechmachtheory.2021.104339.
- [31] M. Ulrich and C. Steger, “Hand-eye calibration of SCARA robots using dual quaternions,” *Pattern Recognition and Image Analysis*, vol. 26, no. 1, pp. 231–239, 2016, doi: 10.1134/S1054661816010272.
- [32] H. Chih-Ching, K. Chin-Hsing, M. Daisuke, and T. Yukio, *Advances in Mechanism and Machine Science*, vol. 73, no. 16. Springer International Publishing, 2019. doi: 10.1007/978-3-030-20131-9.
- [33] P. K. Jamwal, A. Kapsalyamov, S. Hussain, and M. H. Ghayesh, “Performance based design optimization of an intrinsically compliant 6-dof parallel robot,” *Mechanics Based Design of Structures and Machines*, vol. 50, no. 4, pp. 1237–1252, 2022, doi: 10.1080/15397734.2020.1746669.
- [34] P. Li, T. Shu, W. F. Xie, and W. Tian, “Dynamic Visual Servoing of A 6-RSS Parallel Robot Based on Optical CMM,” *Journal of Intelligent and Robotic Systems: Theory and Applications*, vol. 102, no. 2, 2021, doi: 10.1007/s10846-021-01402-5.
- [35] M. Zhang and J. Yan, “A data-driven method for optimizing the energy consumption of industrial robots,” *Journal of Cleaner Production*, vol. 285, p. 124862, 2021, doi: 10.1016/j.jclepro.2020.124862.
- [36] L. F. F. Furtado, E. Villani, L. G. Trabasso, and R. Sutério, “A method to improve the use of 6-dof robots as machine tools,” *International Journal of Advanced Manufacturing Technology*, vol. 92, no. 5–8, pp. 2487–2502, 2017, doi: 10.1007/s00170-017-0336-8.
- [37] P. Bobka, F. Gabriel, and K. Dröder, “Fast and precise pick and place stacking of limp fuel cell components supported by artificial neural networks,” *CIRP Annals*, vol. 69, no. 1, pp. 1–4, 2020, doi: 10.1016/j.cirp.2020.04.103.
- [38] Ş. Yıldırım and S. Savaş, “Design of a Mobile Robot to Work in Hospitals and Trajectory Planning Using Proposed Neural Networks Predictors,” *Lecture Notes in Networks and Systems*, vol. 305, no. 9, pp. 32–45, 2022, doi: 10.1007/978-3-030-83368-8_4.
- [39] B. Belzile, P. K. Eskandary, and J. Angeles, “Workspace Determination and Feedback Control of a Pick-And-Place Parallel Robot: Analysis and Experiments,” *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 40–47, 2020, doi: 10.1109/LRA.2019.2945468.
- [40] Z. A. Karam and E. A. Saeed, “Smartly Control, Interface and Tracking for Pick and Place Robot Based on Multi Sensors and Vision Detection,” *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 1215–1229, 2022, doi: 10.12785/ijcds/120197.
- [41] D. Ionescu *et al.*, “Communication and Control of an Assembly, Disassembly and Repair Flexible Manufacturing Technology on a Mechatronics Line Assisted by an Autonomous Robotic System,” *Inventions*, vol. 7, no. 2, 2022, doi: 10.3390/inventions7020043.
- [42] S. Li, D. Chen, and J. Wang, “A TRAJECTORY TRACKING METHOD OF PARALLEL MANIPULATOR BASED ON KINEMATIC CONTROL ALGORITHM,” *International Journal of Robotics and Automation*, vol. 37, no. 3, pp. 273–279, 2022, doi: 10.2316/J.2022.206-0577.
- [43] A. Singletary, S. Kolathaya, and A. D. Ames, “Safety-Critical Kinematic Control of Robotic Systems,” *IEEE Control Systems Letters*, vol. 6, no. c, pp. 139–144, 2022, doi: 10.1109/LCSYS.2021.3050609.
- [44] P. Boscaroli and D. Richiedei, “Trajectory design for energy savings in redundant robotic cells,” *Robotics*, vol. 8, no. 1, 2019, doi: 10.3390/robotics8010015.
- [45] Q. Xiao, G. Xiang, Y. Chen, Y. Zhu, and S. Dian, “Time-Optimal Trajectory Planning of Flexible Manipulator Moving along Multi-Constraint Continuous Path and Avoiding Obstacles,” *Processes*, vol. 11, no. 1, 2023, doi: 10.3390/pr11010254.
- [46] T. P. Kapusi, T. I. Erdei, G. Husi, and A. Hajdu, “Application of Deep Learning in the Deployment of an Industrial SCARA Machine for Real-Time Object Detection,” *Robotics*, vol. 11, no. 4, 2022, doi: 10.3390/robotics11040069.
- [47] R. Chiba, T. Arai, T. Ueyama, T. Ogata, and J. Ota, “Working environment design for effective palletizing with a 6-DOF manipulator,” *International Journal of Advanced Robotic Systems*, vol. 13, no. 2, pp. 1–8, 2016, doi: 10.5772/62345.

- [48] D. A. Jimenez-Nixon, M. C. Paredes-Sanchez, and A. M. Reyes-Duke, "Design, construction and control of a SCARA robot prototype with 5 DOF," *Proceedings of the 2022 IEEE International Conference on Machine Learning and Applied Network Technologies, ICMLANT 2022*, 2022, doi: 10.1109/ICMLANT56191.2022.9996479.
- [49] M. E. Uk, F. B. Sajjad Ali Shah, M. Soyaslan, and O. Eldogan, "Modeling, control, and simulation of a SCARA PRR-type robot manipulator," *Scientia Iranica*, vol. 27, no. 1, pp. 330–340, 2020, doi: 10.24200/sci.2018.51214.2065.
- [50] A. Bahani, M. E. houssine Ech-Chhibat, H. Samri, and H. A. Elattar, "The Inverse Kinematics Evaluation of 6-DOF Robots in Cooperative Tasks Using Virtual Modeling Design and Artificial Intelligence Tools," *International Journal of Mechanical Engineering and Robotics Research*, pp. 121–130, 2023, doi: 10.18178/IJMERR.12.2.121-130.
- [51] H. Zhu, W. Xu, B. Yu, F. Ding, L. Cheng, and J. Huang, "A Novel Hybrid Algorithm for the Forward Kinematics Problem of 6 DOF Based on Neural Networks," *Sensors*, vol. 22, no. 14, pp. 1–18, 2022, doi: 10.3390/s22145318.
- [52] A. Córdova, P. Quimbiamba, L. J. Segura, L. Escobar, and D. Loza, "Implementation of Collaborative Work Between Two SCARA Robots in a Robotic Cell for Continuous Classification of Products," *Lecture Notes in Electrical Engineering*, vol. 932 LNEE, pp. 97–111, 2022, doi: 10.1007/978-3-031-08288-7_7.
- [53] C. Han, H. Ma, W. Zuo, S. Chen, and X. Zhang, "A general 6-DOF industrial robot arm control system based on Linux and FPGA," *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pp. 1220–1225, 2018, doi: 10.1109/CCDC.2018.8407315.
- [54] P. Ji, C. Li, and F. Ma, "Sliding Mode Control of Manipulator Based on Improved Reaching Law and Sliding Surface," *Mathematics*, vol. 10, no. 11, 2022, doi: 10.3390/math10111935.
- [55] J. Cornejo *et al.*, "Industrial, Collaborative and Mobile Robotics in Latin America: Review of Mechatronic Technologies for Advanced Automation," *Emerging Science Journal*, 2023.

Integrating Transfer Learning and Deep Neural Networks for Accurate Medical Disease Diagnosis from Multi-Modal Data

Dr. Chamandeep Kaur¹, Dr. Abdul Rahman Mohammed Al-Ansari², Dr. Taviti Naidu Gongada³,
Dr. K. Aanandha Saravanan⁴, Divvela Srinivasa Rao⁵, Ricardo Fernando Cosio Borda⁶, R. Manikandan⁷

Lecturer, Department of CS & IT, Jazan University, Saudi Arabia¹

Department of Surgery, Salmania Hospital, Bahrian²

Assistant Professor, Dept. of Operations-GITAM School of Business, Gitam University, Visakhapatnam³

Department of ECE-Vel Tech Rangarajan, Dr. Sagunthala R&D Institute of Science and Technology⁴

Sr. Assistant Professor, Department of AI & DS, Lakireddy Bali Reddy College of Engineering, Mylavaram⁵

Universidad Autónoma de Ica, Peru., Peru⁶

Vel Tech Rangarajan, Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India-600062⁷

Abstract—Effective patient treatment and care depend heavily on accurate disease diagnosis. The availability of multi-modal medical data in recent years, such as genetic profiles, clinical reports, and imaging scans, has created new possibilities for increasing diagnostic precision. However, because of their inherent complexity and variability, analyzing and integrating these varied data types present significant challenges. In order to overcome the difficulties of precise medical disease diagnosis using multi-modal data, this research suggests a novel approach that combines Transfer Learning (TL) and Deep Neural Networks (DNN). An image dataset that included images from various stages of Alzheimer's disease (AD) was collected from kaggle repository. In order to improve the quality of the signals or images for further analysis, a Gaussian filter is applied during the preprocessing stage to smooth out and reduce noise in the input data. The features are then extracted using Gray-Level Co-occurrence Matrix (GLCM). TL makes it possible for the model to use the information gained from previously trained models in other domains, requiring less training time and data. The trained model used in this approach is AlexNet. The classification of the disease is done using DNN. This integrated approach improves diagnostic precision particularly in scenarios with limited data availability. The study assesses the effectiveness of the suggested method for diagnosing AD, focusing on evaluation metrics such as accuracy, precision, miss rate, recall, F1-score, and the Area under the Receiver Operating Characteristic Curve (AUC-ROC). The approach is a promising tool for medical professionals to make more accurate and timely diagnoses, which will ultimately improve patient outcomes and healthcare practices. The results show significant improvements in accuracy (99.32%).

Keywords—Transfer learning; deep neural network; disease diagnosis; multi-modal data; Alexnet; GLCM; DNN; pre-trained model

I. INTRODUCTION

In the fields of machine learning and artificial intelligence, TL is a potent and well-liked technique that aims to use information learned from one task or domain to enhance the performance of another task or domain that is related to it [1]. It is predicated on the notion that knowledge obtained while

resolving one problem can be applied and transferred to resolve another problem that is unrelated but nonetheless similar more effectively [2]. Models are created from scratch for each distinct task using traditional machine learning techniques and lots of labelled data. However, this procedure can be time-consuming, costly in terms of computation, and it may call for a sizable amount of labelled data, which isn't always available. Transfer learning overcomes these constraints by applying previously learned features or representations from a pre-trained model, referred to as the "source task," to a new target task with a smaller dataset. Pre-training and fine-tuning are typically the two essential steps in the TL process. A DNN is trained on a sizable dataset from the source task, such as image recognition on a sizable image dataset, during the pre-training phase [3]. The pre-trained model gains knowledge of broader features and patterns that can be used for a variety of tasks. Numerous industries, including healthcare, speech recognition, computer vision, and NLP, have found extensive use for TL. It has made it possible to create complex models that perform better even when there is a dearth of labelled data. TL encourages knowledge sharing and transfer across tasks by reusing learned representations, resulting in ML models that are more effective and efficient [4].

Alzheimer's disease is a progressive and irreversible neurological condition that primarily affects memory, thought, and behavior in the brain's cognitive regions [5]. It is the most typical cause of dementia, which is a group of brain disorders marked by a decline in memory, communication, and reasoning skills and the inability to perform daily tasks. Usually, a disease takes time to develop and gradually gets worse. People may experience mild memory loss in the early stages, as well as trouble finding words or organizing their thoughts. People with advanced Alzheimer's may experience confusion, mood swings, difficulty solving problems, and a loss of recognition of familiar faces and environments [6]. A vital component of healthcare is accurate disease diagnosis, which enables patients to receive timely and efficient care. In recent years, new opportunities for enhancing patient care and

improving diagnostic accuracy have emerged due to the accessibility of a variety of medical data from various modalities. Data like genetic profiles, clinical reports, medical images, and textual information are included in this [7]. However, because of their complexity, heterogeneity, and the requirement to capture complex relationships between various modalities, analyzing and integrating these multi-modal data sources present significant challenges [8].

This research suggests a novel method that combines TL and DNN for precise medical disease diagnosis from multi-modal data to address these issues [9]. With the ability to transfer knowledge from a source domain to a target domain, transfer learning has become a potent machine learning technique [10]. TL enables the adaptation of learned representations and weights to new datasets, even when the data distributions differ significantly, by utilizing pre-trained models on large-scale datasets [11]. This knowledge transfer improves the ability of models to generalize and makes it easier to make an accurate diagnosis in the target domain [12]. In a number of industries, including computer vision, natural language processing, and healthcare, deep neural networks have achieved remarkable success [13]. These networks provide a strong framework for analyzing multi-modal medical data due to their capacity to learn intricate patterns and relationships from high-dimensional data. Recurrent neural networks (RNNs) capture temporal dependencies in sequential data, while convolutional neural networks (CNNs) are excellent at extracting features from images [14]. Furthermore, attention mechanisms allow the network to concentrate on pertinent data within the multi-modal data, increasing diagnostic precision [15].

There are several benefits to combining TL and DNN when diagnosing medical diseases using multimodal data [16]. First of all, it permits the use of prior knowledge and learned representations from comparable tasks or domains, which can greatly improve the performance of models on small medical datasets. Second, it combines the advantages of various network architectures to enable thorough analysis of multi-modal data by capturing both spatial and temporal dependencies [17]. Last but not least, it offers the possibility of individualized and accurate diagnosis, resulting in enhanced patient outcomes and improved treatment strategies [18]. A branch of ML and AI called "deep learning" focuses on teaching artificial neural networks how to carry out challenging tasks [19]. DNN, which are made up of multiple layers of interconnected nodes (neurons) that process and transform data, are used in this process [20]. These networks can learn and recognize patterns and features from enormous amounts of data because they are built to mimic the structure and operation of the human brain.

The goal of this study is to combine DNN and TL to create a reliable and accurate framework for diagnosing medical diseases [21]. The superiority of the approach over conventional methods and single-modal analyses through extensive tests and evaluations. Additionally, ablation studies are carried out to investigate the influence of various network architectures and TL strategies on diagnostic precision [22].

The key contributions of the paper is,

- The suggested model was trained using an image dataset that was taken from the Kaggle repository and contained images from different stages of AD.
- Before further analysis or feature extraction, the preprocessing stage of the data is where the Gaussian filter is used to smooth the input data and reduce noise.
- The study uses a pre-trained model for transfer learning called AlexNet, which is a well-known DNN architecture.
- The GLCM is used to extract features from the input data, capturing the spatial relationships and occurrence patterns of various pixel intensities in the image, which provides useful texture information for subsequent analysis or classification tasks.
- In order to accurately and efficiently classify diseases, Artificial Neural Networks (ANNs), a type of DNN, are used in the classification task to learn complex patterns and relationships from the input data.
- The research employs a number of evaluation metrics to assess the performance of the proposed approach for AD diagnosis. Accuracy, precision, recall, miss rate, F1-score, and the Area under the Receiver Operating Characteristic Curve (AUC-ROC) are some of these metrics.

The rest of this article is structured as follows: An overview of related research is given in Section II. The problem statement is presented in Section III. The methodology and architecture of our suggested approach are described in Section IV. Section V discusses the findings and subsequent discussion, and Section VI discusses the conclusion.

II. RELATED WORKS

Alzheimer's disease (AD) has become a growing problem among older people [23]. It is crucial for AD treatment to accurately identify mild cognitive impairment in the initial phases of the symptoms. Nevertheless aren't many samples of brain images and they come in a variety of methods, making it very challenging for machines to correctly categorize images of the brain. Through re-transfer training and multi-modal learning, the present research suggests an extremely fine brain image identification method to identifying AD. Diffusion tensor images (DTI) are initially finely classified into four different groups using a throughout its entirety DNN classification system called CNN4AD. Additionally, the re-transfer technique for learning is suggested on the basis of multipurpose theory of learning and is in accordance with the features of the multi-modal brain image data collection. The suggested strategy achieves greater precision with fewer labelled samples for training, according to the experiment's findings. This might aid in a quicker and more precise diagnosis of AD by medical professionals.

AD is a severe and unchangeable dementia of the central nervous system that impairs memory and ability to think [24]. DL algorithms have been proven successful in medical applications in treating this neuro-degenerative illness that results in neurological impairment and mental decline. DL techniques have been discovered to be efficient for duties like recognizing trends in imaging data and helping with diagnosis. When there is a lack of information, which applies an established model to a novel assignment, might prove especially helpful. TL has been used by investigators to successfully identify AD. AD is a severe and unchangeable dementia of the central nervous system that impairs memory and ability to think. DL algorithms have been proven successful in medical applications in treating this neuro-degenerative illness that results in neurological impairment and mental decline. DL techniques have been discovered to be efficient for duties like recognizing trends in imaging data and helping with diagnosis. When there is a lack of information, which applies an established model to a novel assignment, might prove especially helpful. TL has been used by investigators to effectively detect AD.

In modern healthcare environment, diagnostic testing has taken on an important function [25]. Brain cancer, among a particularly deadly disease and the main cause of death worldwide, is a significant area of study in the discipline of healthcare imaging. A rapid and reliable diagnosis made using MRI can enhance the study and outlook of brain tumors. Healthcare visuals needs to be recognized, divided, and categorized in order for automated diagnosis techniques to help physicians in determining the presence of brain tumors. It is crucial for establishing a computerized approach because radiologists find it tedious and prone to errors to manually identify malignancies in the brain. As an outcome, an accurate strategy for identifying and classifying brain tumors is given. The suggested process entails five phases. The borders in the original image are found using a linear contrast enhancement in the initial step. The following step involves creating a unique, 17 layered DNN architecture for segmenting brain tumors. The next step involves training the altered MobileNetV2 design employing for extracting features. The most desirable characteristics were chosen in the following step using an entropy-based regulated technique and a M-SVM. On the information sets from BraTS 2018 and Figshare, the approach that was suggested was tested. An investigation demonstrates that the suggested approach for classifying and detecting brain tumors surpasses existing techniques both qualitatively and in quantitative terms, with accuracy rates of 97.47% and 98.92%, accordingly The XAI technique is then used to clarify the outcome. The suggested approach performed better than existing approaches for identifying and classifying brain tumors. These results show that the suggested method performed better in the context of increased quantitative assessments with better accuracy as well as visual appeal.

After surgical procedures, tumors in breasts patients frequently experience recurring and metastases [26]. For the creation of accuracy therapy, forecasting a person's likelihood of metastatic growth and recurrence is crucial. In the present investigation, histopathological images that were stained with

H&E, medical records, and information concerning gene expression to offer an innovative multi-modal DL forecasting model. To be more precise, DNN to record every image inhibit into a 1D incorporate vector after segmenting tumor spots in H&E into image segments. The probable likelihood of recurrence as well as metastasis for every participant was then predicted by the attention-getting component, which scored every region of the H&E-stained images and paired visual characteristics with a medical and gene transcription information gathered. All 196 cancerous breast specimens with concurrently accessible clinical, expression of genes, and H&E data from the cancer genome database to test the hypothesis. The geographic distributions of the collected data were subsequently maintained among the two databases by centralized collection as the specimens were split into the training and testing sets in a ratio of 7:3. On the evaluation set, the multi-modal model outperformed those that relied merely on H&E image, arranging the information, and medical information, each achieving an AOC value of 0.75. This research could potentially be useful in the clinical setting to determine individuals with breast cancer who are at high risk for responding well to following surgery adjuvant therapy.

A neurodegenerative condition known as Alzheimer's disease (AD), it affects numerous individuals all over the world [27]. Although AD remains one of the most prevalent brain disorders, it can be challenging to identify, and in order to distinguish comparable trends, it needs a classification depiction of its features. Neural networks are commonly used in research to address increasingly difficult issues, including AD detection. Researchers and scientists without specialized expertise in artificial intelligence see those methods as well-understood and even sufficient. Therefore, it is crucial to find a detection technique that is both fully automated and simple for non-AI specialists to utilize. To quickly streamline the creation of neural networks and consequently democratize artificial intelligence, the approach should determine effective settings for the modeling variables. Multi-modal medical image fusion also provides deeper multimodal characteristics and a better capacity for information representation. For more precise diagnostics and more effective therapy, a fusion image is created by combining pertinent and related information from many input images. In order to diagnose Alzheimer's disease, the present research introduces a MultiAz-Net, a novel optimized ensemble-based deep neural network learning model that incorporates heterogeneous data gathered from PET and MRI scans. The study provides an automated method for anticipating the early start of AD according to characteristics identified from the fused data. The suggested structure involves three steps: picture fusion, feature extraction, and classification. A multi-objective optimization technique called the Multi-Objective Grasshopper Optimization technique (MOGOA) is provided to optimize the MultiAz-Net's layers. To do this, the required functions of objectives are enforced and the appropriate values for the proposed variables are looked for. Using the openly accessible Alzheimer neuroimaging dataset, the suggested deep ensemble model was empirically evaluated to complete four tasks for grouping Alzheimer's illness, three binary categorizations, along with a multi-class categorization task.

III. PROBLEM STATEMENT

Limited diagnostic accuracy results from traditional machine learning methods' inability to fully grasp the subtleties found in multimodal medical data. As each modality offers distinct insights into a patient's status, it has become clear that it is necessary to combine different sources of information, such as pictures and clinical data. The accuracy of diagnosing various medical disorders might be greatly increased by creating an integrated framework that smoothly integrates multiple modalities. Additionally, DNNs' performance in image analysis tasks points to their promise in this situation. On the other hand, over fitting and unsatisfactory outcomes frequently arise from training DNNs from scratch on scant medical data. Therefore, there is a need for a technique that efficiently uses transfer learning to adapt pre-trained models to medical diagnosis tasks while maximizing the use of multi-modal data. The development of appropriate data fusion strategies that capture the complimentary information provided by each modality is necessary for the integration of multi-modal data.

To guarantee an effective translation of pre-trained models to the medical domain, where the data distribution may differ greatly from general datasets, transfer learning algorithms must be developed. Choosing the right neural network topologies and optimization techniques to handle the complicated, high-dimensional medical data is also essential. The suggested framework should also take into account the ethical ramifications of using AI in medical diagnosis, including transparency and interpretability. This study's main goal is to provide a novel method for precise medical condition detection utilizing multimodal data that blends transfer learning and deep neural networks. The suggested framework seeks to intelligently leverage the pre-trained knowledge from other domains and efficiently incorporate various medical data sources in order to get beyond the constraints of existing machine learning approaches and increase diagnosis accuracy. The research's ultimate goal is to advance the area of medical diagnostics by enabling more

accurate and early illness identification, which can enhance patient outcomes and make healthcare systems more effective [28].

IV. PROPOSED TL-DNN FRAMEWORK

The methodology involved using an image dataset from Kaggle that contained images from various stages of AD to train the suggested model. Data preprocessing used the Gaussian filter to smooth out and reduce noise before analysis. The pre-trained AlexNet DNN architecture was used with transfer learning. The GLCM was used for feature extraction in order to record spatial relationships and pixel intensity patterns for texture data. ANN were used to extract complex patterns from the input data and classify diseases. Different metrics were also used in performance evaluation. Fig. 1 shows the proposed methodology.

A. Data Collection

The effectiveness of the study would be enhanced by having access to a comprehensive dataset that contains a wide range of disorders, several imaging modalities (such as MRI, CT, and X-ray), and a variety of patient demographics [29]. Such a dataset would enable the examination of the suggested approach's performance across various medical diseases, imaging modalities, and patient groups, providing a more thorough evaluation of the method. It would be possible to evaluate how well the technique generalizes and adjusts to various illness presentations and demographic characteristics using this extensive dataset, which would eventually increase its potential usefulness and relevance in real-world clinical settings. The intended database originated from a publicly accessible Kaggle repository [30]. These MRI images of the brains of individuals who are Very Mildly Demented (VMD), Moderately Demented (MOD), Non-Demented (ND), and Mildly Demented make up this dataset. An image dataset that included images from various stages of AD was used to train the suggested model. Table I shows the overall amount of image samples used as input after enhancement, broken down by class.

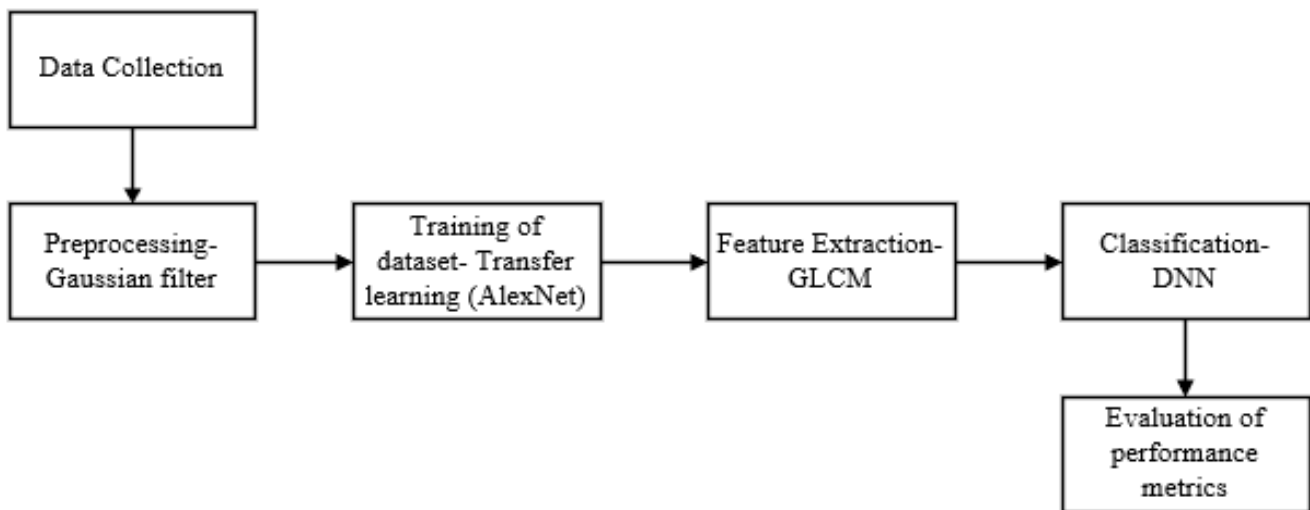


Fig. 1. Proposed methodology.

TABLE I. DATASET PARAMETERS

Mental State	No. of image samples
VMD	1792
MOD	1024
ND	2560
MD	1017

B. Multi-Modal Data Fusion using Autoencoders

AI-driven multi-modal data fusion approaches have evolved as a way to tap into the deep insights contained in these diverse sources. Utilizing autoencoders, a kind of neural networks created expressly for unsupervised feature learning and data reduction, is one well-known strategy in this field. Autoencoders provide a strong foundation for multi-modal data fusion since they are typically used for dimensionality reduction and data reconstruction tasks. They serve as modality-specific encoders in this situation, extracting unique patterns and characteristics from each data source. The models may learn representations that capture modality-specific information while abstracting away noise and unnecessary features by training distinct autoencoders for each modality. In order to merge the encoded representations from various modalities, a shared latent space must be created, and this is where the multi-modal fusion's key challenge resides. The model may learn shared characteristics and relationships that might not be obvious in individual modalities alone by using this joint space as a bridge to enable the integration of modalities. The constraints presented by multi-modal medical data, where complicated interrelationships frequently influence diagnostic findings, are well matched with the flexibility of autoencoders in capturing nuanced connections.

The possibility for higher data quality is one of the intrinsic benefits of utilizing autoencoders for multi-modal fusion. Autoencoders naturally filter out unimportant fluctuations by condensing noisy and high-dimensional input into a latent space, improving the signal-to-noise ratio. This can result in conclusions that are more reliable and generalizable, especially when working with noisy medical data. The clinical relevance and interpretability of multi-modal data fusion employing autoencoders ultimately determines its effectiveness. The combined representations should increase diagnostic precision while also giving doctors useful new information. Making sure that the AI-driven fusion effectively integrates with medical expertise and decision-making involves translating these learnt qualities back into clinical words that can be understood by patients.

C. Preprocessing

Applying a Gaussian filter to an image as part of preprocessing it with a Gaussian function helps to smooth it out and reduce noise. The Gaussian function, a mathematical function that has a bell-shaped curve, serves as the filtering operation's kernel or mask. A Gaussian kernel is used in the

filter's operation, and each pixel is given a weight based on how close to its neighbors it is. The values of the Gaussian function at each point along the kernel are used to calculate the weights. The degree of image smoothing can be altered by modifying the Gaussian filter's parameters, such as the kernel size and standard deviation. Smaller kernel sizes and lower standard deviations preserve finer details while larger kernel sizes and higher standard deviations result in more extensive smoothing. A preprocessed image with less noise, fewer high-frequency details, and a smoother overall appearance is the end result. A common preprocessing method in image processing, Gaussian filtering is especially beneficial for tasks like demising, feature extraction, and improving image quality. Below is the equation for the Gaussian function.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\pi\sigma^2}} \quad (1)$$

Where the standard deviation of the distribution is denoted as σ . The distribution is assumed to have a mean of 0.

D. Employing Transfer Learning for training the Dataset

1) *Pre trained model – AlexNet*: A convolutional network framework that has already been trained is called AlexNet. Transfer learning is the process for employing an approach that has been trained, and it is currently widely employed in applications that employ DL. In the suggested technique, the study employed an updated form of this AlexNet framework. AlexNet is an eight-layer structure with accessible variables, five of which comprise layers of convolution that combine maximum pooling with three layers that are completely interconnected. ReLU is a nonlinear function of activation that is present in every layer. Images obtained from the Pre-processed layers are retrieved by the network inputs layer. Pre-processing, which may be performed in a variety of methods, such as by enhancing specific image characteristics or decreasing the image, is an essential phase in order to produce appropriate datasets. Image scaling is a necessary procedure since images come in a variety of sizes. As a result, images were reduced in dimension to $227 * 227 * 3$, where $227 * 227$ denotes the input images' height and breadth and three indicates the total amount of channels.

The approach incorporates the previously trained CNN network as well as AlexNet, which has a significant influence on contemporary deep learning techniques. After being modified to accommodate the needs, this CNN network was used to feed the preprocessed images to the suggested AlexNet transfer learning system. The resultant of the categorization layer, completely interconnected layer, and softmax layer constitute three substantially adjusted layers in the design that correspond to the issue specification. Fig. 2 depicts the altered network utilized for transfer learning.

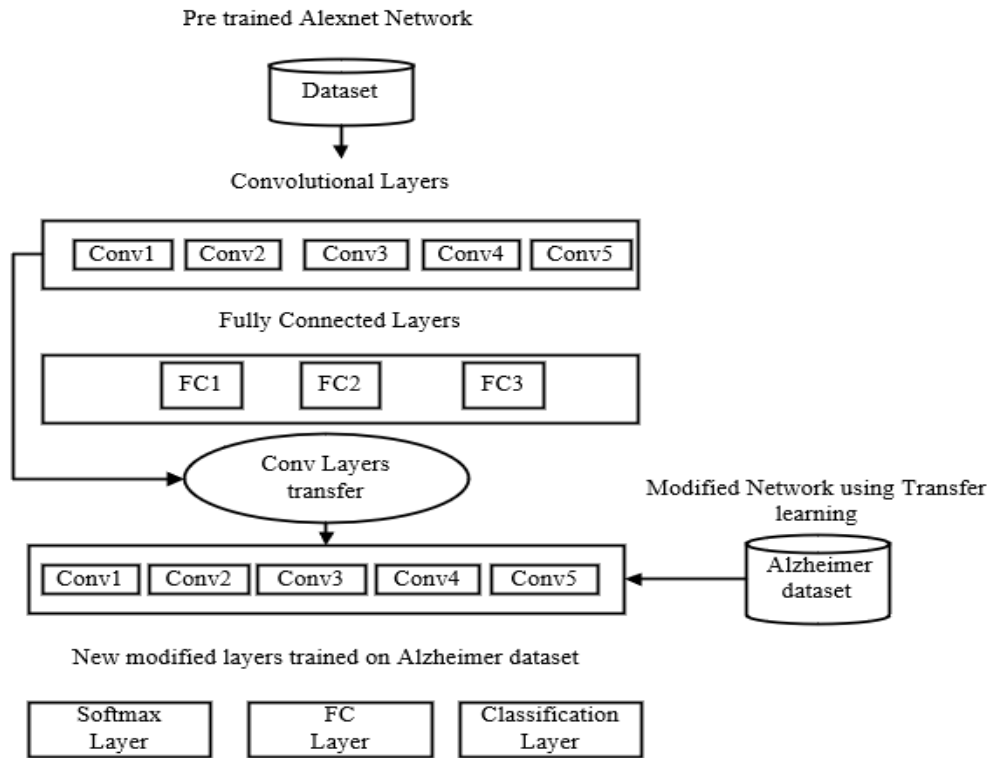


Fig. 2. Pre trained transfer learning model.

Although the subsequent modification layers are learned with Alzheimer datasets, the first five segments of the network, which had been trained employing AlexNet, become unchanged. The remaining three sections are set up to categorize the images into their specific category labels based on the resultant labels assigned to each class. The dimensions of the results, which is made up of several categories, is one of the input variables for completely interconnected layers. A completely linked layer's generated size is equal to the entire quantity of labelled classes. Softmax functions are applied to the data provided using softmax layers. A layer that is completely connected is used for acquiring the class-specific characteristics needed to distinguish across categories, whereas a layer made up of convolutions represents generalized visual characteristics like edge recognition throughout training. As a result, the class-specific characteristics are adjusted for fully linked layers. To categorize images into several classes, the suggested model is trained employing multi-class labelling of Alzheimer's disease.

Numerous variables can be utilized for preparing the system for training, or there are other training choices that can be provided. The remaining three layers of AlexNet—fully interconnected layers, outputs categorization layers and SoftMax layers—are not included in the extraction process since they are not necessary for transfer learning.

The following variables can be utilized as training alternatives: learning rates, the amount of iterations, the rate of validation, and the total amount of epochs.

The different components of CNN are in responsibility for extracting the universal properties from the images, which are referred to as the domain of origin. The resulting learning characteristics can then be applied to determine and categorize a variety of additional operations, such as the identification of Alzheimer's disease. Customized generated models that may be utilized for validation are positioned on clouds. The approach that has been trained evaluates each image and classifies them according to their corresponding categories, which are MD, MOD, VMD and ND. The trained system has precisely the characteristics for image processing that it obtained throughout the training phase. Thus, it was shown that the effectiveness of Alzheimer's disease recognition is influenced by the identification of Alzheimer's disease phases in individuals and the transfer of information from big datasets.

E. Feature Extraction using GLCM

By computing various statistical measures from the GLCM matrix, different facets of the image texture can be captured during feature extraction. These measurements, also referred to as GLCM features or texture features, offer numerical data about the spatial relationships between gray-level values in an image.

1) *Energy*: By adding up the squared values of each component in the GLCM matrix, the energy also known as the angular second moment or uniformity is determined. It gauges the texture of the image's overall intensity or contrast. While a lower energy value reflects a more complex or heterogeneous texture with variations in gray-level values, a higher energy value denotes a more homogeneous texture where the gray-level values are evenly distributed.

$$E = \sum_p \sum_q \{N(p, q)\}^2 \tag{2}$$

Where the images are denoted as N, and the image's squares with grey levels are labeled as (p, q).

2) *Contrast*: In terms of the variations in their grey levels, contrast quantifies the intensity contrast between pixel pairs. It displays how much local variation or abrupt texture transitions there are. Low contrast values imply a more uniform or smooth texture, while high contrast values suggest significant differences between adjacent pixel pairs.

$$C = \sum_{y=0}^{I_q} y^2 \left\{ \sum_{p=1}^{I_q} \sum_{q=1}^{I_q} N(p, q) \right\} \tag{3}$$

I stand for the grayscale of the images, N for the images, and (p, q) for the square of an image's grayscale.

3) *Correlation*: Correlation measures how closely the linear associations between the gray-level values of adjacent pixels are related to one another. It shows how the values of the grey levels vary consistently or predictably between adjacent pixels. A stronger or more nonlinear relationship between the gray-level values is indicated by a higher correlation value than by a lower correlation value.

$$C_o = \frac{\sum_p \sum_q (p, q)N(p, q) - \mu_u \mu_v}{\sigma_u \sigma_v} \tag{4}$$

In the images, the values of mean, as well as standard deviation, are μ_u , μ_v , σ_u , and σ_v are characterized as row and column.

4) *Entropy*: The amount of information or uncertainty related to the texture is measured by entropy. It displays the distribution of gray-level values among adjacent pixel pairs. A texture with a higher entropy value is more varied or heterogeneous, with dispersed and unpredictable gray-level values. A lower entropy value, on the other hand, denotes a more regular or uniform texture, where the values of the grey levels are concentrated or predictable.

$$En = - \sum_p \sum_q N(p, q) \log(N(p, q)) \tag{5}$$

F. Classification using DNN

DNN have an intricate network model and adhere to the same structure as standard ANN. It aids in the development of

models and the clear definition of complex structure. It has 'n' layers that are hidden that analyze information obtained from the layer preceding it, which is referred to as the initial layer. Following every moment, the rate of errors of the input information is going to be gradually lowered by changing the weights for every node, back promoting the network's structure and continuing until it achieves improved outcomes. In the input layer, any amount of the inputs can be designated as input nodes. In order to intensify the conditioning analyze, DNN typically has multiple nodes than the data it receives from the layer. As distinct nodes for output in the layer that produces the results, any amount of generates can be stated. The total quantity of points in the data for input and output, bias, developing rate, starting weights for modification, the amount of hidden layers, the total number of nodes in each hidden layer, and prevent the requirements for stopping the operation of the times are considered to be the variables that are utilized by the DNN. In order to prevent network leads to from being invalidated discrimination value is typically set to 1 in any neural network design. Additionally, the rate of learning is set to 0.15 by standard and is later arbitrarily impacted through trial and error to produce different results compared to the equation. The system can determine the beginning weights of the nodes at arbitrary, modify it throughout replication by determining its error rate, and modify it frequently after every phase. The amount of inputs and the dimension of the information determine the amount of secret layers as well as node locations in each hidden layer. Both the system reaches the ideal amount of periods or the anticipated outcome from the approach to learning is realized is referred to as the network's removal state. It requires a greater amount of resources to train the computational framework if there are more sections and node locations in the framework.

1) *Artificial neural network*: ANN can be used to classify cases of AD. An example of a ML model is an ANN, which takes inspiration from the design and operation of neural networks in biology. They are made up of interrelated "neurons," or nodes, organized in levels. Every neuron takes in information, uses a stimulus to create an output, and subsequently sends the result to the neural layer below. Fig. 3 shows the architecture of ANN.

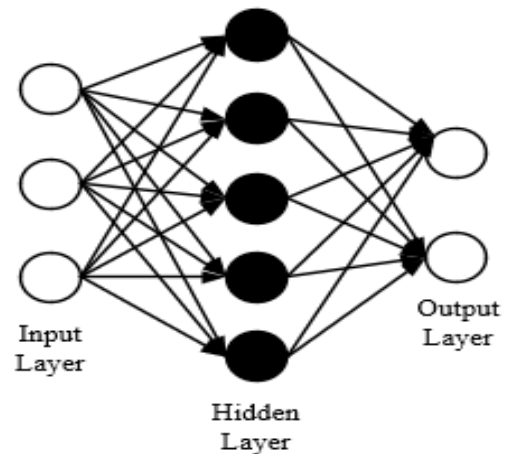


Fig. 3. Architecture of ANN.

To avoid the biased impact caused by the input data due to the different ranges, each input distinctive, $b_i; 1 \leq i \leq 5$, had been normalized inside the same range $[y, a]$. Let's say an additional range is defined as $[y, x]$. Let b_{max} and b_{min} represent the upper and lower bounds of the concept's natural range. The distinctive values b_i within the range of values $[y, x]$ could be normalized using the formula that follows symbolized as b'_i .

$$b'_i = \frac{(y-x)(b_i-b_{min})}{b_{max}-b_{min}} + y \quad (6)$$

A symmetrical function has x and y parameters of 0.1 and 0.9, respectively, whereas a tangent hyperbola function has x and y coordinates of 0.9 and 0.9 in both cases. The symmetrical function is shown below.

$$f(b) = [1 + \exp(-b)]^{-1} \quad (7)$$

The highest and lowest values of the hyperbola's tangent activating function were limited to $[-1, 1]$, while the highest and lowest values of the symmetrical function had been limited to $[0, 1]$. The tangential hyperbolic function of activation is shown in Eq. (3).

$$f(b) = [\exp b - \exp(-b)]/[\exp(b) + \exp(b)] \quad (8)$$

This limits the network's expected output for symmetrical and hyperbolic angular functions of activation, respectively, to $[0, 1]$ and $[1, 1]$. On the contrary hand, the system's outputs don't accurately reflect the data's true worth. The output value has to be changed to its actual value, b'_d using the following,

$$b'_d = b_{min} + \left[\frac{b_d - y}{x - b} \right] \quad (9)$$

Equations 10 and 11 calculate the generalization error using the summation squared errors (SSE) and regression analysis error (R2):

$$\text{summation squared error} = \sum(b'_d - b_t)^2 \quad (10)$$

$$\text{Regression analysis} = 1 - [\sum(b'_d - b_t)^2 / \sum(b'_d - b)^2] \quad (11)$$

Where b is the measurement value of the evaluation sequence t and b_t is the average of the data collected.

V. RESULTS AND DISCUSSION

The methodology involved training the suggested model using an image dataset from Kaggle that included images from various stages of AD. Before analysis, noise in the data was smoothed out and reduced using the Gaussian filter. Transfer learning was used in conjunction with the pre-trained AlexNet DNN architecture. In order to capture spatial relationships and pixel intensity patterns for texture data, the GLCM was used for feature extraction. ANN were used to classify diseases and extract intricate patterns from the input data. Additionally, various metrics were applied when assessing performance.

A. Accuracy

The system model's overall performance is assessed using accuracy. Essentially, it is the notion that each encounter will be accurately predicted. Accuracy is provided in equation (12),

$$\text{Accuracy} = \frac{T_{Pos} + T_{Neg}}{T_{Pos} + T_{Neg} + F_{Pos} + F_{Neg}} \quad (12)$$

TABLE II. COMPARISON OF ACCURACY

Method	Accuracy (%)
Inception V4 [31]	73.75
Landmark based feature extraction [31]	79.02
ADDTLA [31]	91.7
TL-DNN	99.32

The accuracy of the suggested TL-DNN and the existing methods is shown in Table II. Due to its accuracy value of 99.32%, the suggested method is determined to be more effective than the others. The accuracy is shown in Fig. 4.

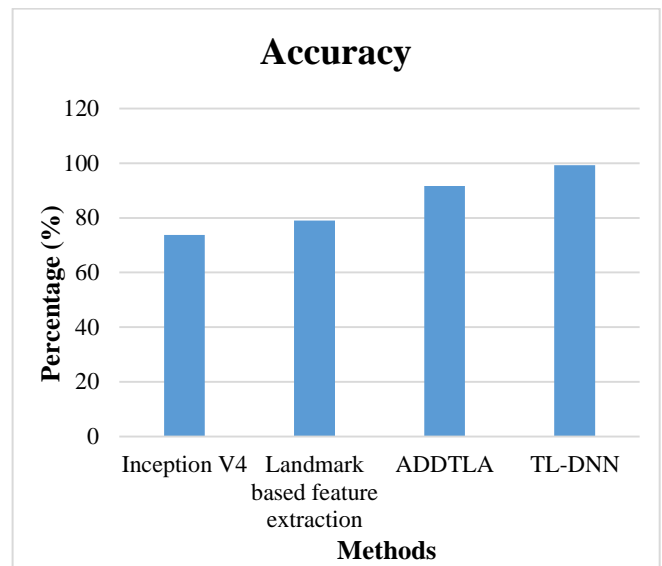


Fig. 4. Comparison of accuracy.

B. Precision

Besides to being correct, precision also refers to how closely two or more calculations resemble one another. The relationship between accuracy and precision demonstrates how repeatedly a finding can be made. Equation (13) can be used to calculate precision.

$$P = \frac{T_{Pos}}{T_{Pos} + F_{Pos}} \quad (13)$$

C. Recall

Recall is the proportion of all pertinent results that the methods were effectively sorted. The ratio among the true positive and false negative values is used to calculate the appropriate positive for those numbers. It is referred to in equation (14).

$$R = \frac{T_{Pos}}{T_{Pos} + F_{Neg}} \quad (14)$$

D. F1-Score

The F1-Score formula combines recall and accuracy. Precision and recall are used to calculate the F1-Score that is given in equation (15).

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (15)$$

TABLE III. COMPARISON OF PERFORMANCE METRICS

Method	Precision (%)	Recall (%)	F1-score (%)
DNN [32]	97	97	97
SMO [32]	94.1	93.9	96.3
LDA [32]	95.7	95.5	95.5
KNN [32]	89.2	86.4	86.6
TL-DNN	98.12	98	98.5

Table III compares the precision, recall, and F1-score of current methods with the suggested TL-DNN. The precision of the recommended TL-DNN is higher than that of the other approaches. In Fig. 5, it is shown.

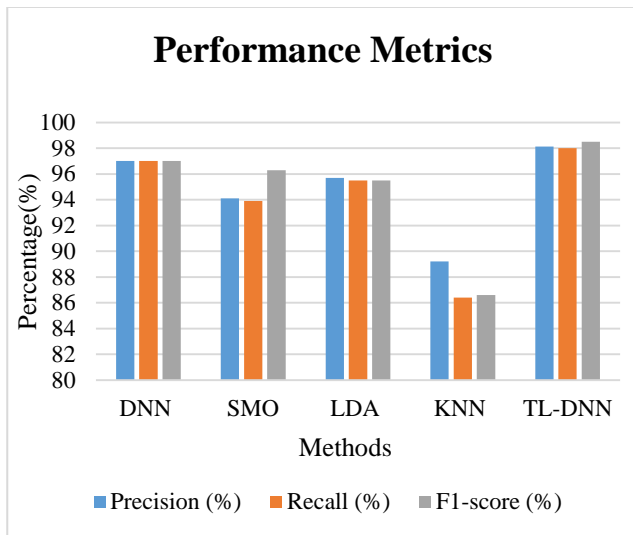


Fig. 5. Performance Metrics.

E. Miss Rate

A binary classification model's effectiveness is measured by the miss rate, also referred to as the false negative rate or Type II error rate. It calculates the percentage of positive instances that the classifier actually classifies as negative.

$$MissRate = \frac{F_{Neg}}{T_{Pos} + F_{Neg}} \quad (16)$$

The miss rate of the proposed TL-DNN is compared with the miss rate of other methods, which is given in Table IV. The miss rate of TL-DNN is 5.1% which lower than other methods. It is represented in Fig. 6.

F. Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The area under the ROC curve is known as the AUC. It evaluates the classifier's overall performance in separating the

two classes while taking into account all potential classification thresholds. A classifier's performance is graphically represented by the ROC curve. Fig. 7 depicts the AUC-ROC.

TABLE IV. COMPARISON OF MISS RATE

Methods	Miss Rate (%)
Inception V4 [31]	26.25
Landmark based feature extraction [31]	20.98
ADDTLA [31]	8.3
TL-DNN	5.1

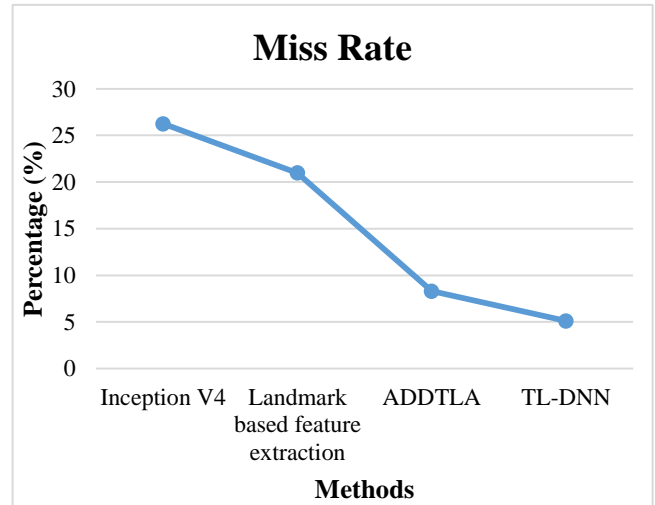


Fig. 6. Comparison of miss rate.

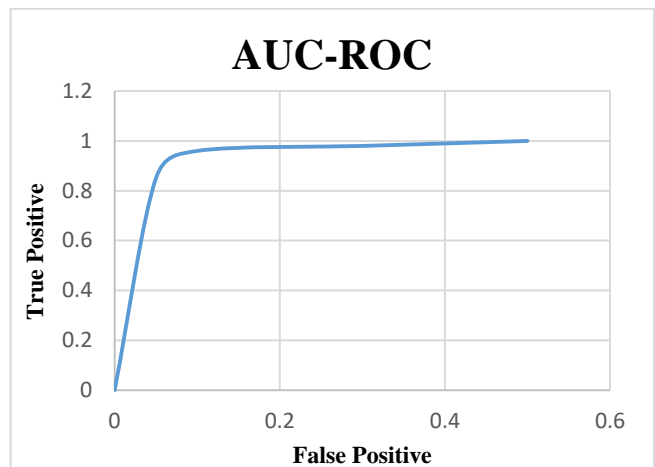


Fig. 7. Area under the receiver operating characteristic curve.

G. Discussion

The end result discusses how various metrics can be used to evaluate a system model's effectiveness when performing binary classification tasks. Evaluation of the accuracy, precision, miss rate, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are the main considerations. Accuracy is a binary classification metric that shows how frequently the model predicts correctly. The accuracy of various methods is compared in the table and

figure, with TL-DNN having the highest accuracy (99.32%). Less false positives are implied by higher precision. Comparing precision, Table III shows that TL-DNN has the highest precision (98.12%). The miss rate for the TL-DNN is the lowest (5.1%). The F1-Score and recall values achieved by TL-DNN are 98% and 98.5%, respectively. The outcome offers a thorough analysis of the model's performance using a variety of metrics, enabling a thorough evaluation of its efficiency in the binary classification task. The outcomes demonstrate that the TL-DNN method performs remarkably across a variety of evaluation metrics.

VI. CONCLUSION

For accurate medical disease diagnosis using multimodal data, the combination of TL and DNN shows to be a highly effective and promising approach. This study effectively illustrates the potential to improve diagnostic precision by utilizing information from pre-trained models and imaging data sources. Even with little training data, the model can generalize to different medical specialties and diseases by using transfer learning. Due to its robustness and dependability in diagnosing medical conditions, this integrated approach has a lot of potential for use in the real world. The model's achieved high generalization and accuracy highlight its practical value in aiding medical professionals in making prompt and accurate diagnoses. This strategy can result in better treatment choices, better patient outcomes, and possibly lower healthcare costs by accurately diagnosing diseases. Despite these noteworthy developments, there are still some issues that need to be resolved. Maintaining the model's performance in real-world medical settings requires careful consideration of data quality and mitigating potential biases. This study offers insightful information into the field of medical disease diagnosis and demonstrates the enormous potential of fusing DNN and TL. The deployment of extremely precise and trustworthy diagnostic tools that will revolutionize medical procedures and have a big impact on patient care by overcoming obstacles and further improving the strategy. Innovative and ethical solutions to benefit patients and the healthcare industry as a whole will be developed as the field of artificial intelligence in healthcare develops as a result of ongoing research and collaboration between AI experts and healthcare professionals. The legal implications of utilizing AI in medical diagnosis, including transparency, interpretability, and possible biases, are not adequately covered in the research. The investigation of other designs that can perhaps produce superior results may be constrained by the employment of a specific neural network architecture (AlexNet).

REFERENCES

- [1] P. Xi, C. Shu, and R. Goubran, "A Unified Deep Learning Framework for Multi-Modal Multi-Dimensional Data," in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, 2019, pp. 1–6.
- [2] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 361, pp. 185–195, 2019.
- [3] G. Lee, K. Nho, B. Kang, K.-A. Sohn, and D. Kim, "Predicting Alzheimer's disease progression using multi-modal deep learning approach," *Scientific reports*, vol. 9, no. 1, p. 1952, 2019.
- [4] G. Sreeja and O. Saraniya, "Image fusion through deep convolutional neural network," in *Deep learning and parallel computing environment for bioengineering systems*, Elsevier, 2019, pp. 37–52.
- [5] Z. Xiang *et al.*, "Self-supervised multi-modal fusion network for multi-modal thyroid ultrasound image diagnosis," *Computers in Biology and Medicine*, vol. 150, p. 106164, Nov. 2022, doi: 10.1016/j.compbiomed.2022.106164.
- [6] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/su15031906.
- [7] L. Xu, M. Bennamoun, F. Boussaid, S. An, and F. Sohel, "Coral classification using densenet and cross-modality transfer learning," in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2019, pp. 1–8.
- [8] U. Evci, V. Dumoulin, H. Larochelle, and M. C. Mozer, "Head2Toe: Utilizing Intermediate Representations for Better Transfer Learning," in *Proceedings of the 39th International Conference on Machine Learning*, PMLR, Jun. 2022, pp. 6009–6033. Accessed: Aug. 22, 2023. [Online]. Available: <https://proceedings.mlr.press/v162/evci22a.html>
- [9] "Preparing for the next pandemic via transfer learning from existing diseases with hierarchical multi-modal BERT: a study on COVID-19 outcome prediction | Scientific Reports." <https://www.nature.com/articles/s41598-022-13072-w> (accessed Jul. 13, 2023).
- [10] Q. Zhu, N. Yuan, J. Huang, X. Hao, and D. Zhang, "Multi-modal AD classification via self-paced latent correlation analysis," *Neurocomputing*, vol. 355, pp. 143–154, 2019.
- [11] J. He, C. Zhou, X. Ma, T. Berg-Kirkpatrick, and G. Neubig, "Towards a Unified View of Parameter-Efficient Transfer Learning." arXiv, Feb. 02, 2022. doi: 10.48550/arXiv.2110.04366.
- [12] N. E. Benzebouchi, N. Azizi, A. S. Ashour, N. Dey, and R. S. Sherratt, "Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 6, pp. 841–874, 2019.
- [13] A. S. K. Reddy *et al.*, "Multi-modal fusion of deep transfer learning based COVID-19 diagnosis and classification using chest x-ray images," *Multimed Tools Appl*, vol. 82, no. 8, pp. 12653–12677, Mar. 2023, doi: 10.1007/s11042-022-13739-6.
- [14] Y. Pan, M. Liu, C. Lian, Y. Xia, and D. Shen, "Disease-image specific generative adversarial network for brain disease diagnosis with incomplete multi-modal neuroimages," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22*, Springer, 2019, pp. 137–145.
- [15] J. Hu *et al.*, "Towards accurate and robust multi-modal medical image registration using contrastive metric learning," *IEEE Access*, vol. 7, pp. 132816–132827, 2019.
- [16] J. Kalajdjieski *et al.*, "Air Pollution Prediction with Multi-Modal Data and Deep Neural Networks," *Remote Sensing*, vol. 12, no. 24, Art. no. 24, Jan. 2020, doi: 10.3390/rs12244142.
- [17] W. Liu *et al.*, "Research on medical data feature extraction and intelligent recognition technology based on convolutional neural network," *IEEE Access*, vol. 7, pp. 150157–150167, 2019.
- [18] "Sensors | Free Full-Text | Learning for a Robot: Deep Reinforcement Learning, Imitation Learning, Transfer Learning." <https://www.mdpi.com/1424-8220/21/4/1278> (accessed Aug. 22, 2023).
- [19] K. Oh, Y.-C. Chung, K. W. Kim, W.-S. Kim, and I.-S. Oh, "Classification and visualization of Alzheimer's disease using volumetric convolutional neural network and transfer learning," *Scientific Reports*, vol. 9, no. 1, p. 18150, 2019.
- [20] "RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning | Radiology: Artificial Intelligence." <https://pubs.rsna.org/doi/full/10.1148/ryai.210315> (accessed Aug. 22, 2023).
- [21] O. Pelka, F. Nensa, and C. M. Friedrich, "Branding-fusion of meta data and musculoskeletal radiographs for multi-modal diagnostic recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.

- [22] S. Ismail, B. Ismail, I. Siddiqi, and U. Akram, "PCG classification through spectrogram using transfer learning," *Biomedical Signal Processing and Control*, vol. 79, p. 104075, Jan. 2023, doi: 10.1016/j.bspc.2022.104075.
- [23] M. Fang, Z. Jin, F. Qin, Y. Peng, C. Jiang, and Z. Pan, "Re-transfer learning and multi-modal learning assisted early diagnosis of Alzheimer's disease," *Multimed Tools Appl*, vol. 81, no. 20, pp. 29159–29175, Aug. 2022, doi: 10.1007/s11042-022-11911-6.
- [24] S. K. G. U., A. P. R., and S. K. G., "Detecting Alzheimer's Disease using Multi-Modal Data: An Approach Combining Transfer Learning and Ensemble Learning," in *2023 International Conference on Control, Communication and Computing (ICCC)*, May 2023, pp. 1–6. doi: 10.1109/ICCC57789.2023.10165454.
- [25] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-Modal Brain Tumor Detection Using Deep Neural Network and Multiclass SVM," *Medicina*, vol. 58, no. 8, Art. no. 8, Aug. 2022, doi: 10.3390/medicina58081090.
- [26] Y. Yao *et al.*, "ICSDA: a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data," *Briefings in Bioinformatics*, vol. 23, no. 6, p. bbac448, Nov. 2022, doi: 10.1093/bib/bbac448.
- [27] W. N. Ismail, F. R. P. P., and M. A. S. Ali, "A Meta-Heuristic Multi-Objective Optimization Method for Alzheimer's Disease Detection Based on Multi-Modal Data," *Mathematics*, vol. 11, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/math11040957.
- [28] C.-Y. Wee *et al.*, "Cortical graph neural network for AD and MCI diagnosis and transfer learning across populations," *NeuroImage: Clinical*, vol. 23, p. 101929, 2019.
- [29] A. El-Said *et al.*, "Assessing the Impact of Demographic Factors on Presenting Conditions or Complaints Among Internal Medicine Patients in an Underserved Population in Central Florida," *Cureus*, vol. 14, no. 8, p. e27811, doi: 10.7759/cureus.27811.
- [30] "Alzheimer's Dataset (4 class of Images)." <https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images> (accessed Jul. 14, 2023).
- [31] T. M. Ghazal *et al.*, "Alzheimer Disease Detection Empowered with Transfer Learning," *Computers, Materials & Continua*, vol. 70, no. 3, pp. 5005–5019, 2022, doi: 10.32604/cmc.2022.020866.
- [32] H. Mohsen, E.-S. A. El-Dahshan, E.-S. M. El-Horbaty, and A.-B. M. Salem, "Classification using deep learning neural networks for brain tumors," *Future Computing and Informatics Journal*, vol. 3, no. 1, pp. 68–71, Jun. 2018, doi: 10.1016/j.fcij.2017.12.001.

An Integrated Instrument for Measuring Science, Technology, Engineering, and Mathematics: Digital Educational Game Acceptance and Player Experience

Husna Hafiza R. Azami¹, Roslina Ibrahim², Suraya Masrom³,
Rasimah Che Mohd Yusoff⁴, Suraya Yaacob⁵

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia^{1, 2, 4, 5}
Computing Sciences Studies, College of Computing, Informatics and Media,
Universiti Teknologi MARA Perak Branch, Tapah Campus, 35400 Malaysia³

Abstract—Digital educational games (DEGs) are effective learning tools for subjects related to science, technology, engineering, and mathematics (STEM), yet they are still not widely used among students. Existing instruments typically assess player experience (PX) and acceptance separately, even though both are essential DEG evaluations that can be merged and analyzed concurrently in a thorough manner. This study, therefore, proposes an integrated instrument called DEGAPX that combines fundamental technology acceptance factors with a broad range of PX criteria. The proposed instrument can be used by educators and game designers in the selection and development of DEGs that satisfy the needs of target users. This article describes the process of developing the scale instrument and validating it through two rounds of expert judgment and among students after using three DEGs related to STEM. The proposed instrument, which comprised 15 constructs measured by 67 items, was proven to be reliable and valid.

Keywords—Game; education; acceptance; experience; stem

I. INTRODUCTION

The complexity of the modern world, which necessitates that the workforce be prepared with the knowledge and abilities to tackle cross-disciplinary problems, is argued to make science, technology, engineering, and mathematics (STEM) education necessary [1]. In contrast to the growing demand, there have been fewer STEM graduates in many countries [2]. STEM-related degrees accounted for six out of ten of the educational programs, with the highest percentage of dropouts among 44,406 students at 20 Malaysian public institutions [3]. A universal concern in education is learner's lack of enthusiasm in STEM fields.

The use of modern pedagogical approaches such as educational gaming technologies has been advocated by academics as a way to improve students' interest, engagement, and performance in learning [4], [5]. Among 408 university students in Malaysia, more than 60% of them prefer using online games to supplement their studies because they believe games can make their learning more enjoyable [6].

Moreover, it has been demonstrated that game-based learning applications may improve learners' understanding, interconnection, and exploration of scientific and mathematical concepts. This instructional tool can give students

opportunities for experiential learning that let them apply the concepts learned in the classroom to actual situations while also developing their creativity, critical thinking, and problem-solving abilities [7]. Furthermore, the COVID-19 pandemic highlights the importance of educational technologies like DEGs as a preparation for an unpredictable future that may require remote instruction. These technologies enable the acquisition of knowledge and skills outside of the classroom [8], [9].

Digital educational games (DEGs) are online or offline applications in electronic devices that integrate fun and educational elements [10]. This technology is anticipated to be widely used and accepted because current learners are people who have grown up with a heavy reliance on the internet and other information technologies such as digital games [11]. It has been shown that around 80% of all internet users worldwide between the ages of 16 and 44 play video games in the year 2022 [12]. Therefore, DEGs are useful and pertinent technological applications that can support students' learning.

Although DEGs offer great promise and are a good fit for today's learners, they have not yet received widespread acceptance. Game developers face challenges in creating successful, well-received DEGs since they are more difficult to design than commercial entertainment games due to the need to serve both educational and recreational goals [13]. Numerous DEGs have fallen short of the expectations placed upon them in terms of learning or enjoyment outcomes [14].

With the growth of DEGs, more research is required to assist game developers, instructors, and policymakers in the design, development, selection, and implementation of DEGs that satisfy students' preferences as the key target users. When an information system is widely adopted by the intended audience, it can be considered successful [13]. Although there has been various researches that investigate the predictors of DEG acceptance, a systematic literature review shows that most of them primarily concentrate on technological acceptance viewpoints, which are insufficient to fully comprehend students' preferences for DEGs [15].

Given the distinctive and complex features of DEGs, a variety of player experience (PX) factors that contribute to players' pleasure should be incorporated into the acceptance

model and instrument since it can affect users' decisions to utilize the technology [16]. Existing studies normally evaluate DEG acceptance and PX separately [19]. Performance expectancy, effort expectancy, and social influence are prevalent characteristics of technology acceptance [17], [18]. By incorporating PX factors, the assessment of acceptance can be expanded to include crucial DEG design features such as enjoyment, relevant content, feedback, and challenges.

Hence, the purpose of this research is to present an inclusive DEG acceptance instrument called DEGAPX that integrates PX factors. The instrument is crucial and helpful in designing and developing potentially popular DEGs. The instrument can be used by DEG developers to adapt the design that suits users' needs to boost the likelihood of DEG success. The instrument can also be utilized as a guide by instructors when identifying which DEGs could appeal to their students. The subsequent sections of this paper comprise a literature review, methods, results, and discussion, followed by a conclusion in the last section.

II. LITERATURE REVIEW

A. Digital Educational Game Acceptance Factors

According to Dillon and Morris [20], acceptance refers to the willingness of individuals to use an information system that has been created for them. Current approaches to technology development and adoption have considered the requirement for predictive measures of users' likely usage in order to judge the success of a technology. The technology may be deemed unsuccessful if the majority of the population rejects it. By analyzing the data obtained for the measurement scales in questionnaires, previous studies have frequently been able to forecast and explain why people want to use a particular technology [17], [18].

One of the most well-known theories for predicting human behavior when it comes to probable technology acceptance or rejection is the unified theory of acceptance and use of technology (UTAUT). The theory combines eight other renowned theories such as the technology acceptance model (TAM) for explaining and forecasting technology usage [18]. Three independent variables in UTAUT namely performance expectancy, effort expectancy, and social influence predict behavioral intention that affects use behavior.

UTAUT had previously been used to investigate what criteria led Malaysian university students to choose DEGs for learning programming [21]. In place of social influence and facilitating condition variables from the original UTAUT model, the research added attitude, self-efficacy, enjoyment, and anxiety. With the exception of self-efficacy and anxiety, all other independent variables were seen to have a substantial impact on students' intention to utilize the DEG. The use behavior construct was left out with justification that DEG was still a relatively new technology in the area where the study took place and the students were still unfamiliar with it.

Wan et al. [22] look into what influences undergraduate students' acceptance of six digital board games and their ability for independent learning. The instrument was derived from the integration of UTAUT with flow theory as well as the motivated strategies for learning questionnaire (MSLQ). To

identify the causes of primary school students' intentions to continue using mobile games for learning mathematics, another study also made an effort to combine multiple theories such as flow theory and the game-based learning model [23]. Nevertheless, the only variable from technology acceptance was the ease of use.

B. Digital Educational Game Player Experience Factors

According to ISO 9241-210:210 (clause 2.15), user experience (UX) is defined as the way a person feels and behaves after using or anticipating the use of a system, product, or service. For digital educational games (DEGs), UX is sometimes referred to as player experience (PX). Assessment of PX broadens game usability evaluation by focusing on meaningful and enjoyable experiences of users rather than only getting rid of technological barriers [24].

PX is a crucial factor in determining how long a person will play DEGs which will determine the DEG's success [25]. When a game has a strong PX, consumers are more likely to play it frequently, stay engaged for a longer duration, and recommend it to others [26]. Despite the wide variety of learning games available today, some of them are unattractive to consumers and have low retention rates, as consumers get bored after a few gaming sessions [27]. The absence of elements that can improve PX may be one of the causes. Consumers may lose interest when their playing experience falls short of their expectations. Therefore, it is critical to include PX in DEG evaluation.

There have been many different methods for evaluating PX, including questionnaire scales, field studies in real-world settings, lab studies, and online studies where participants can be anonymous [28]. Since studies on the variables that determine user acceptance generally use measurable constructs evaluated using Likert scale questionnaires, prior literature that assessed PX in DEG using a similar method was reviewed. The PX factors derived from those studies can then be analyzed in the same manner as the acceptance factors without any problems.

One of the most popular measurement scales used by researchers to gauge how consumers feel while playing games is the game experience questionnaire (GEQ) [29]. Seven different factors are measured by the questionnaire's items, including the positive affect (enjoyment), negative affect, flow, challenge, tension, as well as sensor and imaginative immersion, which relates to a rich gaming experience, beautiful design, and engaging game plot. However, there were no learning-related factors in the GEQ scale.

Nagalingam et al. [31] proposed one of the recent instruments to thoroughly assess PX in digital learning games called the educational game experience (EDUGX). The instrument which had been reviewed by experts and tested among 273 computer science diploma students, had demonstrated content validity, internal consistency, convergent validity, and discriminant validity. There were six components to the instrument including immersion, usability, flow, player context, and learnability, each of which was broken down into a number of sub-components. The immersion component gauges how invested individuals feel in the game they play,

whereas usability measures the extent to which players found the game to be effective and satisfying. On the other hand, learnability represents the educational aspect of the game; player context pertains to the user background and the social interaction that the game supports, while flow indicates the state of total focus.

III. METHODS

A. Instrument Development

The development of the DEGAPX scale instrument for this research follows the guideline by DeVellis [32], where the first step entails reviewing theories from past studies relevant to the research objective. Since this research intends to propose a comprehensive instrument for measuring the acceptance of DEGs with the integration of PX components, constructs that evaluate DEG acceptance and PX in prior studies were identified from two separate systematic literature reviews [15].

1) *Determination of constructs*: Performance expectancy, effort expectancy, and social influence are the constructs in UTAUT by Venkatesh et al. [18] which had been proven by many studies to have an impact on students' intention to use DEGs [22], [33], [34]. Hence, this study chooses these three constructs together with behavioral intention from UTAUT to represent the core technology acceptance constructs in the proposed DEGAPX instrument.

From the review of PX constructs used by scholars [27], those proposed by Nagalingam et al. [31] in the EDUGX framework can be considered for this research since they were developed after taking into account diverse PX criteria in other studies. However, some adjustments were made.

For instance, EDUGX [31] and the frameworks in other existing research [35], [36] used a control factor to assess the extent of freedom felt by players over the game menu, character movement, actions, and strategies. Since DEG is an instructional tool, this study decides to transform the construct into a learning control construct that focuses on players' perceptions of control over their learning recovery, problem-solving approaches, and ability to choose the game content that they want to learn and the difficulty level.

Apart from that, under the game usability component in EDUGX [31], operability was defined as the game performance including its accessibility, ease of use, and lack of technological glitches, while understandability was defined as the game's messages, functions, inputs, and outputs being simple to understand. On the other side, the game system sub-component indicated how well a gaming gadget operated in terms of being simple and comfortable to use. These revealed that operability, game system, and understandability from the EDUGX framework [31] and effort expectancy from the UTAUT model [18] were comparable. Therefore, this research chose to use the effort expectancy construct to cover the game usability measurement items in EDUGX. In addition, knowledge improvement under the learnability component of EDUGX is similar to the UTAUT model's performance expectancy construct, which measured the game's capacity to enhance students' learning performance.

As a result, 14 constructs were considered for this study, as shown in Fig. 1, including performance expectancy, effort expectancy, social influence, and behavioral intention from UTAUT [18], as well as learning relevance, attractiveness, enjoyment, challenge, clear goal, learning control, social interaction, feedback, concentration, and immersion modified from EDUGX [31] for player experience measurement.

Construct	Definition
Player experience	
Learning relevance	The degree to which the game instructions are appropriate with user's learning goals, previous knowledge and preferred way of learning.
Attractiveness	The level of player attraction to the game as a result of its sensory elements, such as audio and visual.
Enjoyment	The degree to which an individual perceives the activity of using the game is enjoyable.
Challenge	The degree to which users believe the game is difficult enough and appropriate for their skill level.
Clear goal	The extent to which a player perceives the goals are clearly presented.
Learning control	The extent to which users believe they have control over their learning in the game.
Social interaction	The extent to which the game supports and encourages social connection and interaction.
Feedback	The degree that individuals perceive the game provides feedback on progression and accomplishment.
Concentration	The ability of the game to deliver stimuli that will pique players' attention and encourage player's focus.
Immersion	A state in which players believe they are actively participating in the content of the game and completely involved in the game world.
Technology acceptance	
Performance expectancy	The extent to which individuals perceive that using the game will help improve their performance in STEM learning and skills.
Effort expectancy	The extent of ease related to the use of the game.
Social influence	The extent to which individuals perceive that important others believe DEGs should be used.
Behavioral intention	Degree of an individual's intention to use DEGs related to STEM.

Fig. 1. Definition of the 14 constructs in the DEGAPX instrument that integrates technology acceptance and player experience measurement.

2) *Determination of items and measurement format*: After a thorough examination of the validated items used in prior studies, a scale which consisted of 87 items in total was generated to measure the constructs in the DEGAPX instrument. Some items were modified from previous research to suit the context of this study. New items were also proposed, including those under the performance expectancy construct that measure the perception of students on their STEM learning and skill improvement through the DEG.

For the format of responses, students were required to indicate their agreement on each questionnaire item using the

five-point Likert scale with 1 (strongly disagree), 2 (disagree), 3 (not sure), 4 (agree), and 5 (strongly agree). The instrument questions were prepared in Malay, as well as in English language as the respondent's first and second language, respectively. The instrument had been proofread by an English lecturer with a master's degree in Teaching English as a Second Language (TESL), and the Malay translation had also been examined by the Malaysian Institute of Translation and Books (ITBM) to ensure accurate translation.

B. Content Validation through Expert Judgment

For evaluating a new or updated measurement instrument, the establishment of content validity is a crucial first step before performing other validation techniques [37]. Based on the opinions of subject-matter experts, the content validation procedure enables researchers to acquire data on the relevancy, clarity, and comprehensiveness of an instrument.

The instrument for this research was scrutinized by experts in the field pertinent to this research, as shown in Table I. Two rounds of expert review were conducted, following Polit et al. [38] and Tojib [37]. The documents for the expert panel were prepared following the detailed guideline by Elangovan & Sundaravel [39] to ensure that the experts understand what is expected of them and to facilitate the validation process.

TABLE I. EXPERTS PROFILE

Expert	Field of Expertise	Years of Experience
E1	Game-based learning, creative content	20
E2	Game design, educational technology, visual informatics	19
E3	Educational games design and evaluation, acceptance and use of information system, usability, user experience	15
E4	Human computer interaction, science, technology, engineering, and mathematics	15
E5	Game-based learning, e-learning, learning technologies, gamification, augmented reality, virtual reality games	15
E6	Technology acceptance, multimedia in education, augmented reality, science, technology, engineering, and mathematics	15
E7	User experience in educational games	8
E8	Mobile application and games development	6
E9	Information system, technology acceptance and adoption	4

The first round involved nine experts, and the second round was conducted with three of the experts to validate the refined instrument. These numbers of experts are within the suggested range by Polit et al. [38]. The selection criteria for the content experts were those who hold a Ph.D. qualification and actively conduct research in the field of interest or have experience developing DEGs [37].

The content validity index (CVI) is a reliable approach to judge whether the content of a new or revised scale is valid. Another commonly used method for measuring content validity is the content validity ratio (CVR) by Lawshe [40]. The goal of the content validation for this research was to ascertain whether any item needed to be revised or eliminated

based on the CVR value, the CVI value of individual items (I-CVI), as well as whether more items were required in order to fully explore the construct in light of expert opinion from the comment section [38], [40].

The CVR threshold is influenced by the number of experts. For nine panel of experts, items that achieve the CVR value of 0.78 and above can be retained, while the rest can be considered for elimination [40]. Similarly, items with I-CVI values are higher than 0.78 which shows that the scale has excellent content validity, whereas values below 0.78 indicate that the items need to be revised or eliminated. I-CVI can be calculated by dividing the number of experts who gave a three or four rating on a four-point relevance scale by the total number of experts [41].

Low CVI values could indicate that the operationalization of the underlying construct in the items was not good, or information and directions given to the experts were insufficient, or the experts themselves were biased or inadequately skilled [38]. Hence, a lot of effort was put to create high-quality items as well as to choose a strong panel of expert judges. CVI for the overall scale (S-CVI/AVE) can then be determined. While 0.80 is the minimum acceptable value for S-CVI/AVE, 0.90 or above is advised for a scale to be deemed to have great content validity [38].

C. Instrument Testing among Target Respondents

It is important to conduct a pilot study among target respondents to determine whether a specific research instrument is appropriate for use without any errors or shortcomings before it can be employed in a larger scale research. The reliability and validity of the questionnaire items and how well respondents understood the items need to be judged during this stage [42].

The research sample consists of 14 years old students from a public school in Terengganu, one of the states in Malaysia. Approval from the Ministry of Education (MOE) Malaysia and the Terengganu State Education Department was requested before the data collection. Following that, a meeting was held with the school principal to obtain permission and discuss the schedule. All students were given a form for parental or guardian consent.

The research objectives were briefly explained to the participants including the games that they need to play and evaluate, as well as the confidentiality and anonymity of their feedback. They were reminded of the significance of this study, the necessity of reading each survey question thoroughly before responding and to avoid providing incomplete or straight-lining (identical) responses.

Three DEGs were chosen for testing, based on the DEGAPX instrument, such as having relevant content, clear goals, attractive features, and ability to improve students' learning and skills pertinent to STEM. The first DEG decided for this research is a simulation game called Poly Bridge 2 (<https://www.polybridge2.com/>) by Dry Cactus that lets players learn the foundations of bridge design. Players need to construct bridges that work well under specific conditions using the materials provided.

The second DEG in this study, RoboCo (<https://roboco.co/>) by Filament Games, requires students to design robots that can complete a range of tasks. Students can practice their coding skills using Python language when automating their robots. The two games, Poly Bridge and RoboCo, can help players become more creative as well as better at problem-solving and design-thinking, which are important skills in STEM. On the other hand, the third game, Moonbase Alpha (<https://www.nasa.gov/offices/education/programs/national/ltf/games/moonbasealpha/index.html>) was published by NASA and free-to-play. Students can play alone or collaborate in a team with other players to repair equipment and resume oxygen production at the moon outpost. The screenshots of the three games were shown in Fig. 2.

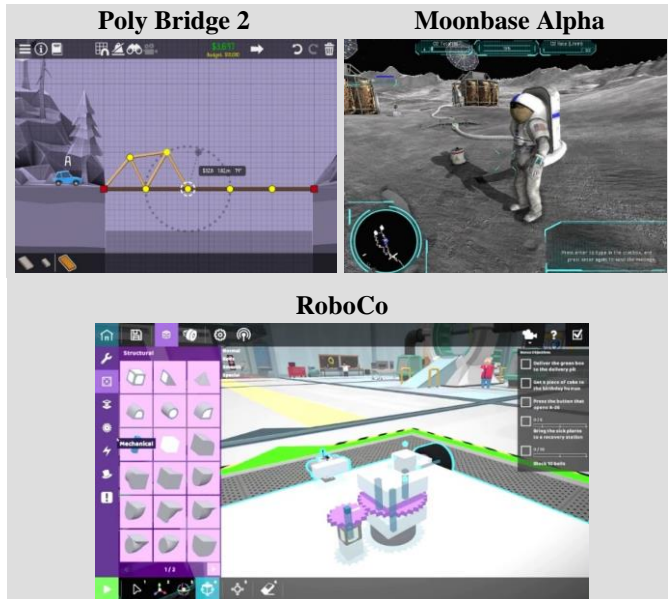


Fig. 2. The images of the three STEM DEGs used in this research.

While some students played all three games, others engaged in one or two games only. They need to answer a self-administered paper-based questionnaire after playing the game for about 90 minutes in the school computer laboratory. There were 281 valid responses obtained from students. RoboCo received 87 responses, Moonbase Alpha received 93 responses, and Poly Bridge 2 received 101 responses.

D. Reliability and Validity Assessment from Instrument Testing

Using SmartPLS 3, reliability was examined based on outer loading of items, Cronbach’s alpha (α), and composite reliability (CR). Outer loadings denote the proportion of the item variance that is explained by the construct, while α and CR assess the intercorrelation between items.

On the other hand, average variance extracted (AVE) can be used to evaluate convergent validity, which refers to the strength of a positive correlation between items. For discriminant validity, heterotrait-monotrait (HTMT) can determine how distinct one construct is from other constructs. Table II showcases the criteria by Hair et al. [43].

Exploratory factor analysis (EFA) with principal component extraction and varimax rotation can also be carried out in IBM SPSS Statistics 27 software to assess construct validity. The data set is appropriate for factor analysis if the Keyser-Meyer-Olkin (KMO) value surpasses 0.5 with a significant Bartlett’s test of sphericity result below 0.05.

TABLE II. RELIABILITY AND VALIDITY ASSESSMENT CRITERIA BY HAIR ET AL. [43]

Criteria	Guidelines
Outer loading	- Remove items with loadings less than 0.4 - Retain items if their loading exceeds 0.7. - Remove items with loadings between 0.4 and 0.7 if AVE and CR can be satisfied.
Cronbach’s alpha (α)	- More than 0.70 is satisfactory. - Between 0.6 and 0.7 is acceptable.
Composite reliability (CR)	- Must exceed 0.70.
Average variance extracted (AVE)	- Must exceed 0.5.
Heterotrait-Monotrait (HTMT)	- Should be less than 0.9

IV. RESULTS

This section describes the results from the expert judgment and instrument testing of the DEGAPX instrument.

A. Content Validity

In terms of the constructs, all experts remarked them to be adequate in representing the whole aspect of DEG, and that none need to be added or removed. Thus, the proposed 14 constructs remained.

In light of first round of content validation results by nine experts for the items used to measure each construct, the CVI of the entire scale (S-CVI/AVE) was found to be 0.97, which passed the minimum acceptability limit. Five items did not meet the 0.78 content validity ratio (CVR) cut off, so they were regarded as weak in representing the particular constructs and can be eliminated. Some experts had issues with the items, and while certain research used them to measure the construct, their inclusion was not deemed vital since they were not included in other studies.

Among 87 items in the initial pool, 65 of them can be accepted without any necessary changes as all of them displayed excellent content validity based on the I-CVI and CVR value and were deemed relevant and clear by the experts. 17 items that also exceeded the 0.78 threshold for I-CVI and CVR, however, need to be refined according to experts’ comments and ratings on clarity. The improvement included rewording, rephrasing, elaborating, and providing examples to improve the comprehensibility of the items. From the experts’ comments, the proposed questionnaire items were deemed adequate, with no additional items needed.

The modified scale underwent a second round of validation with the help of experts E2, E3, and E6. They were one of those who provided a lot of input in the preliminary round. With a 1.00 average scale content validity (S-CVI/AVE) score, the findings showed agreement on all 84 items in the revised DEGAPX instrument, with no additional adjustments being recommended.

B. Reliability, Convergent Validity and Discriminant Validity

During the pilot test, none of the students had any trouble understanding the survey questions. When 281 survey responses obtained were analyzed in SmartPLS 3.0, 17 items need to be removed to satisfy the criteria in Table II. Apart from that, the performance expectancy construct also needs to be separated into two categories based on EFA rotated component matrix to achieve the required value of AVE. As a result, the DEGAPX instrument with 15 constructs measured by 67 items achieved the requirement for reliability and validity, as displayed in Table III. The 67 items are listed in Table IV, with their outer loadings above 0.40.

The suitability for factor analysis was demonstrated from the 0.875 value of KMO and a significant result of Bartlett’s test (n=281; $\chi^2 = 11,997.707$; d.f. = 3,486; p < 0.001). Next, the relationship between the constructs can be investigated to determine the significant predictors of STEM DEG acceptance.

TABLE III. RELIABILITY AND VALIDITY ASSESSMENT RESULTS

Construct	Items	α (>0.6)	CR (>0.7)	AVE (>0.5)
Attractiveness (AT)	5	0.808	0.866	0.566
Challenge (CH)	4	0.711	0.820	0.533
Clear goal (CG)	3	0.688	0.827	0.615
Concentration (CN)	4	0.674	0.802	0.507
Effort expectancy (EE)	7	0.842	0.881	0.514
Enjoyment (EJ)	5	0.784	0.854	0.544
Feedback (FB)	4	0.743	0.833	0.559
Immersion (IM)	5	0.757	0.836	0.506
Behavioral intention (IN)	5	0.754	0.835	0.505
Learning control (LC)	3	0.606	0.793	0.563
Learning performance expectancy (LE)	5	0.749	0.833	0.503
Learning relevance (LR)	4	0.728	0.830	0.553
Skill performance expectancy (SE)	4	0.729	0.831	0.553
Social influence (SF)	4	0.675	0.803	0.507
Social interaction (ST)	5	0.841	0.886	0.609

TABLE IV. DEGAPX INSTRUMENT CONSISTING OF 67 ITEMS

Code	Item Statement	Loading
Learning Relevance Construct		
LR1	DEG content is relevant to my learning need.	0.851
LR2	The way the DEG works suit my way of learning.	0.779

LR3	The DEG content is connected to the other knowledge I already had.	0.652
LR4	Most of the gaming activities are related to the learning task in the DEG.	0.678
Attractiveness Construct		
AT1	I like the general appearance of the DEG.	0.778
AT2	I am attracted to the DEG as a whole.	0.774
AT3	Generally, I find the DEG to be visually appealing.	0.781
AT4	The DEG design is attractive.	0.776
AT5	I like the graphic used in the DEG.	0.642
Enjoyment Construct		
EJ1	I think the DEG is enjoyable.	0.836
EJ2	There were moments when I had fun playing the DEG.	0.795
EJ3	I find the DEG interesting.	0.782
EJ4	Something happened during the DEG playing session that made me smile.	0.651
EJ5	The DEG does not become repetitive or boring as it progresses.	0.592
Challenge Construct		
CH1	My skill gradually improves through the course of overcoming the challenges in the DEG.	0.760
CH2	The difficulty level of challenges increases as my skills improved.	0.707
CH3	The DEG provides new challenges at an appropriate pace.	0.770
CH4	The DEG provides different levels or types of challenges, according to player’s preference.	0.679
Clear Goal Construct		
CG1	Overall game goals are presented in the beginning of the DEG.	0.797
CG2	The intermediate game goals or sub-goals are mostly presented at appropriate times.	0.775
CG3	The game goals are generally clear.	0.780
Learning Control Construct		
LC1	I feel a sense of control over the actions that I take to solve the problems or to achieve better results in the DEG.	0.794
LC2	I feel a sense of control over the strategies that I use to solve the problems or to achieve better results in the DEG.	0.806
LC3	The DEG supports my recovery from errors or mistakes.	0.639
Social Interaction Construct		
SF1	I am able to interact with other people such as other players or friends or online community when playing the DEG.	0.774
SF2	The DEG makes me interact with other people such as for getting help or sharing information.	0.821
SF3	I like to play the DEG with other people.	0.780
SF4	I am able to play the DEG with other players if I choose to.	0.737
SF5	I would enjoy the social interaction through the DEG.	0.786
Feedback Construct		
FB1	I receive feedback on my game progress.	0.855
FB2	I receive immediate feedback on my actions in the DEG.	0.780

FB3	The DEG notifies me immediately when there are new tasks.	0.634
FB4	The DEG notifies me immediately when there are new events.	0.702
Concentration Construct		
CN1	The DEG provides content that stimulates my attention.	0.773
CN2	The DEG provides various stimuli to maintain my attention.	0.818
CN3	Generally, I am not distracted from tasks that the player should concentrate on.	0.592
CN4	I am not burdened with unrelated tasks.	0.637
Immersion Construct		
IM1	I can become less aware of my surroundings if I play the DEG for a long time.	0.643
IM2	The DEG can make me temporarily forget worries about everyday life.	0.788
IM3	I think the DEG can sometimes make me not notice the time passes when playing.	0.761
IM4	I feel emotionally involved in the DEG.	0.633
IM5	I think the DEG can make me spend more time playing than my initial plan.	0.720
Learning Performance Expectancy Construct		
LE1	I would find the DEG useful in my study.	0.709
LE2	Using the DEG would enable me to learn the related subject or concept more quickly.	0.768
LE3	Learning through the DEG would help me to understand the related subject or concept better.	0.784
LE4	The DEG can help me relate the knowledge learnt to real world situations.	0.550
LE5	The DEG would allow me to relate knowledge from multiple learning subjects.	0.710
Skill Performance Expectancy Construct		
SE1	The DEG can help me apply knowledge or skills to situations or practices related to technology or engineering.	0.685
SE2	The DEG can improve my skill in problem-solving.	0.742
SE3	The DEG can improve my creativity skills.	0.722
SE4	The DEG can increase my ability to design, test, and evaluate solutions.	0.819
Effort Expectancy Construct		
EE1	It is easy to learn the related subject or concept or skill through the DEG.	0.656
EE2	I find the DEG easy to use.	0.780
EE3	Learning to use the DEG is easy for me.	0.725
EE4	I think it will be easy for me to become skillful at using the DEG.	0.686
EE5	The interaction with the DEG is clear and understandable.	0.695
EE6	The DEG rules are generally clear and understandable.	0.722
EE7	The DEG instructions are mostly clear and understandable.	0.747
Social Influence Construct		
SF1	People who are important to me think that I should use DEG.	0.654
SF2	I think my school will support the use of DEG.	0.614
SF3	I think my friend or classmate will support the use of DEG.	0.757

SF4	My friend or classmate thinks playing DEG is a good idea.	0.808
Behavioral Intention Construct		
IN1	I intend to use DEG related to STEM in the future.	0.611
IN2	I predict I would use DEG related to STEM in the future.	0.656
IN3	I am interested to play the DEG again.	0.761
IN4	I plan to use the DEG to expand my learning or improve my skill.	0.716
IN5	I am willing to play the DEG frequently.	0.793

V. DISCUSSION

This section discusses the validated 15 constructs of the DEGAPX instrument and the items used to measure them.

A. Learning Relevance

This construct was altered from Luyt et al. [44] and Sideris and Xinogalos [45] with some extensions, where learning relevance was measured not only by players' perceptions of how well the game content corresponds to their existing knowledge and learning needs but also by how closely the game activities matched the learning tasks in the DEG and players' learning styles. Students in this study mostly participated in quiz learning games, which are effective educational interventions for drill-and-practice activities and receiving immediate performance feedback on knowledge learned in classrooms. Because of that, the different learning objectives and approaches of the three STEM DEGs in this study may have an impact on students' perceptions of the learning relevance of the games.

Students may perceive the STEM DEGs have a little amount of instructional content relevant to their prior knowledge since the games place more emphasis on the application of knowledge to solve real-world problems. In addition, students might not find the games meet their educational needs unless they already have a keen interest in pursuing careers pertinent to the games. Nonetheless, students might find the games suit their learning styles and offer appropriate activities, which results in items LR2 and LR4 being added. The four items proposed, which had been proven to be reliable and valid can gauge how players feel about the DEGs' learning relevance from various angles.

B. Attractiveness

This construct was judged based on the overall appearance and design of a DEG, particularly its virtual aesthetics, as derived from Phan [30] and Tao [46]. One of the experts raised the possibility of other multimedia forms that can fall under this attractiveness construct, but the proposed items were deemed sufficient for this research.

There are many different types of DEG multimedia, such as texts, images, and animations, which would result in too many items if they were measured separately and could cause respondent fatigue when answering the survey. Future research that evaluates DEG with fewer constructs can include more items for measuring specific features of DEGs that users can find appealing.

Attractiveness had been demonstrated to have a significant correlation with students' enjoyment when playing DEGs [46]. The suggested items are beneficial for developers to determine the attraction level of their game prototype designed for teaching and learning purposes.

C. Enjoyment

All experts agreed that the five items in Table IV were pertinent for assessing the degree of enjoyment felt by players. Several studies have revealed that students' willingness to play DEGs is significantly influenced by their level of enjoyment [13], [47], [48]. Hence, this construct is one of the most important criteria for a successful DEG.

D. Challenge

The four items proposed were found reliable and valid for representing the challenge construct. This construct is intended for ensuring a DEG provides suitable challenges for players' skill level, which are neither too easy, making players bored, nor too difficult, causing players distress [49]. RoboCo and Poly Bridge 2 games offer a variety of challenges, where players can only access the next challenge after completing the one before it. While some students enjoy demanding games, others might favor easy, relaxing games that do not require much mental effort. Thus, games that allow users to select their preferred difficulty level might appeal to a wide range of consumers.

E. Clear Goal

This research offers three items, as presented in Table IV, for assessing the goal clarity of a DEG as adapted from the validated items in prior research [30]. The items were all regarded as relevant by the experts and can influence students' willingness to play learning games [22].

F. Learning Control

It is believed that players' learning performance can be enhanced when they have control over their learning in the game. Hence, the learning control construct put forth in this study embodies the assistance provided by a game for players to learn from their mistakes as well as the freedom to select their preferred course of action and problem-solving tactics.

G. Social Interaction

The items for this construct were modified from Phan and Keebler [30] and G Petri et al. [14] to measure the extent to which students believe the game promotes social connection, whether it be for knowledge sharing or help-seeking. Additionally, this construct gauges how much students think the game allowed them to play with other players if they want to and how much they enjoy the interaction.

Social interaction is a wonderful game design element that can boost students' enjoyment and motivate them to play a game repeatedly to engage with other players through communication, cooperation, and competition in the game.

Not all games have a multi-player feature, but even without it, social interaction can still be encouraged when players can communicate with other people, such as when getting help or sharing information through an online chat room, discussion forums, or game-based learning activities in classrooms.

Hence, the proposed five items listed in Table IV are generic enough and adequate to measure players' perception of social interaction through a DEG.

H. Feedback

Playing a DEG will be more pleasurable if it offers consumers immediate feedback and informs them of their progress, achievements, failures, and new tasks. The four items in Table IV, which had been constructed based on Nagalingam et al. [31] and Fu et al. [35], were proven reliable and valid.

I. Concentration

The four items for measuring the extent of concentration perceived by students when playing a DEG were developed based on previous research [22], [35]. The items evaluate how much players believe a DEG does not burden them with unnecessary duties, does not take their attention away from activities they should be concentrating on, and provide a variety of stimuli to keep players interested. Concentration was established as a significant determinant of students' instructional computer games acceptance [13]. Thus, it is a crucial factor to consider when developing DEGs.

J. Immersion

Immersion had been demonstrated to have a substantial impact on students' intention to continue using mobile learning games in a previous study [23]. Hence, game creators should design DEGs with fun, challenging activities and interesting features that can make players feel immersed. Items presented in Table IV can be used in the assessment of a DEG since they appropriately reflect the immersion criteria.

K. Performance Expectancy

Due to the focus of this study being STEM DEGs, the survey items from prior research [50] were expanded to incorporate performance expectancy from two categories including knowledge and skill improvement. DEGs for STEM must not only facilitate and improve students' comprehension and performance in learning certain concepts or topics, but also fostering their abilities pertinent to STEM like designing, creativity, and problem-solving.

Given that performance expectancy is one of the most important factors in predicting whether or not students will accept DEGs [51], it is imperative to ensure that any DEG being developed for STEM will benefit students in some way beyond simply enhancing their academic performance. Games' potential should be utilized to provide students with learning opportunities that go beyond a straightforward replication of what is taught from the pedagogical tools traditionally used in classrooms [52].

All items suggested for representing performance expectancy were deemed relevant by experts. The items can be utilized by game developers when designing games intended to enhance students' learning performance and skill development.

L. Effort Expectancy

This construct measures the degree of ease associated with using DEG. Past research typically represents this construct with questionnaire items that gauge how simple it is to learn how to use a game and acquire knowledge or skills through it

[46], [48]. The scope of the items employed in existing literatures when measuring effort expectancy was widened to include game usability components linked to user-friendliness, such as unambiguous rules and instructions.

Since many studies have shown that effort expectancy is a significant predictor of students' intention to use DEGs [15], game developers can consider integrating various game design elements that encourage ease of playing and assess target users' perception using the items shared in Table IV.

M. Social Influence

The survey questions used in existing research [18], [34] that typically measure social influence using items like SF1 in Table IV were expanded to include specific people that may have an influence on students, such as their school, friends, and classmates. Past research had displayed strong association between social influence and students' intention to play DEGs [22][33]. Therefore, positive peer perception and school support towards the use of DEGs may encourage students to use them. Game designers can also consider utilizing community forums, social media pages, and advertisements that improve the game impression and social influence.

N. Behavioral Intention

The sample students of this study had never played the three STEM DEGs, namely Moonbase Alpha, RoboCo, and Poly Bridge 2, before the instrument testing. They were allowed to play the games for about 90 minutes before completing the questionnaire. The opening hour of a game is known to be crucial for hooking and enticing players to keep playing and recommend it to others. A lot of DEGs had been abandoned for the reason that they were not captivating enough to hold players' interest. The initial user experience can have an impact on retention and the possibility that the player will suggest the game to other people.

Hence, five items were proposed to measure students' interest to play a DEG again and frequently, as well as their willingness to use DEGs related to STEM in the future. From the ratings and comments by experts, all five items were found to be relevant and clear for measuring the early acceptance of STEM DEGs among students.

VI. CONCLUSION

This paper describes the development of a scale instrument called DEGAPX and its evaluation through expert judgment and instrument testing. The instrument integrates technology acceptance and player experience (PX) factors for studying and understanding students' perception of digital educational games (DEGs) associated with science, technology, engineering, and mathematics (STEM).

Among the 15 constructs suggested in the DEGAPX instrument, five of them, namely learning performance expectancy, skill performance expectancy, effort expectancy, social influence, and behavioral intention, were derived from the unified theory of acceptance and use of technology (UTAUT). The other ten constructs include learning relevance, attractiveness, enjoyment, challenge, clear goal, learning control, social interaction, feedback, concentration, and immersion.

The instrument with 67 items was found to be reliable and valid after going through two rounds of expert judgment and being tested among 14 years old students after they played three DEGs applicable to STEM, including Poly Bridge 2, Moonbase Alpha, and RoboCo.

The proposed DEGAPX instrument can enrich existing literature on DEG acceptance and PX, especially for STEM education. Despite the various studies available on DEG acceptance, most of them concentrated largely on common technology acceptance factors without thoroughly taking into account the PX elements that are crucial in the design of successful DEGs. Prior research typically evaluates acceptance and PX separately, even though both evaluations can be combined and measured simultaneously in a comprehensive manner. Game developers can utilize the instrument proposed in this research for designing promising DEGs that have the potential to be widely received by students.

For future work, the proposed instrument will be further analyzed using the partial least squares structural equation modeling (PLS-SEM) method to study the relationship between the constructs and figure out the significant determinants of DEG acceptance. This research contains a few limitations. First, because the research sample consisted of 14-year-old Malaysian secondary two students, other research can improve the generalizability by widening the scope to include students from different education levels and learning institutions. This study also emphasizes students' evaluation of DEGs. Future research may look into investigating the perception of other educational stakeholders, including parents, teachers, school administrators, and policymakers.

ACKNOWLEDGMENT

The authors thank the Ministry of Higher Education Malaysia for the research opportunity and fund provided under the Fundamental Research Grant Scheme (grant number FRGS/1/2020/ICT10/UTM/02/2) and Universiti Teknologi Malaysia for the support given.

REFERENCES

- [1] A. Kiray and M. Shelley, *Research Highlights in STEM Education*. USA: ISRES Publishing, International Society for Research in Education and Science (ISRES), 2018. Available online: https://www.isres.org/research-highlights-in-stem-education-7-b.html#_Yvwo5i5ByxU (accessed on 20 March 2023).
- [2] E. Mellander and P. Lind, "Recruitment to STEM studies: The roles of curriculum reforms, flexibility of choice, and attitudes," *Rev. Educ.*, vol. 9, no. 2, pp. 357–398, 2021, doi: 10.1002/rev3.3262.
- [3] O. Wooi Leng, N. Vaghefi, N. Kar Yong, and Y. Jo-Yee, "Student's Choice of STEM Study in Secondary and Tertiary Education in Penang," 2020. Available online: <https://penanginstitute.org/publications/books-and-reports/students-choice-of-stem-study-in-secondary-and-tertiary-education-in-penang/> (accessed on 22 March 2023).
- [4] I. Araújo and A. A. Carvalho, "Enablers and Difficulties in the Implementation of Gamification: A Case Study with Teachers," *Educ. Sci.*, vol. 12, no. 3, 2022, doi: 10.3390/educsci12030191.
- [5] J. K. Acosta-Medina, M. L. Torres-Barreto, and A. F. Cárdenas-Parga, "Students' preference for the use of gamification in virtual learning environments," *Australas. J. Educ. Technol.*, vol. 37, no. 4, pp. 145–158, 2021, doi: <https://doi.org/10.14742/ajet.6512>.
- [6] H. H. Razami and R. Ibrahim, "Distance Education during COVID-19 Pandemic: The Perceptions and Preference of University Students in

- Malaysia Towards Online Learning,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 4, pp. 118–126, 2021, doi: 10.14569/IJACSA.2021.0120416.
- [7] Z. Akimkhanova, K. Turekhanova, and G. P. Karwasz, “Interactive Games and Plays in Teaching Physics and Astronomy,” *Educ. Sci.*, vol. 13, no. 4, pp. 1–22, 2023, doi: 10.3390/educsci13040393.
- [8] K. Cook-Chennault, I. Villanueva Alarcón, and G. Jacob, “Usefulness of Digital Serious Games in Engineering for Diverse Undergraduate Students,” *Educ. Sci.*, vol. 12, no. 1, 2022, doi: 10.3390/EDUCSCI12010027.
- [9] F. Dahalan, N. Alias, and M. S. N. Shaharom, *Gamification and Game Based Learning for Vocational Education and Training: A Systematic Literature Review*, no. 0123456789. Springer US, 2023. doi: 10.1007/s10639-022-11548-w.
- [10] N. Whitton, *Learning with Digital Games: A Practical Guide to Engaging Students in Higher Education*. Routledge, New York, 2010. Available online: <https://www.routledge.com/Learning-with-Digital-Games-A-Practical-Guide-to-Engaging-Students-in-Higher/Whitton/p/book/9780415997751> (accessed on 22 March 2023).
- [11] S. Aslan and O. Balci, “GAMED: digital educational game development methodology,” *Simul. Trans. Soc. Model. Simul. Int.*, vol. 91, no. 4, pp. 307–319, 2015, doi: 10.1177/0037549715572673.
- [12] J. Clement, “Share of internet users worldwide who play video games on any device as of 3rd quarter 2022, by age group and gender,” *Statista*, 2023. Available online: <https://www.statista.com/statistics/326420/console-gamers-gender/> (accessed on 9 March 2023).
- [13] Y. Huang, “Exploring students’ acceptance of educational computer games from the perspective of learning strategy,” *Australas. J. Educ. Technol.*, vol. 35, no. 3, pp. 132–149, 2019, doi: <https://doi.org/10.14742/ajet.3330>.
- [14] G. Petri, C. G. von Wangenheim, and A. F. Borgatto, “A Large-Scale Evaluation of a Model for the Evaluation of Games for Teaching Software Engineering,” in *2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering Education and Training Track (ICSE-SEET)*, 2017, pp. 180–189. doi: 10.1109/ICSE-SEET.2017.11.
- [15] H. Hafiza Razami and R. Ibrahim, “Models and Constructs to Predict Students’ Digital Educational Games Acceptance: A Systematic Literature Review,” *Telemat. Informatics*, vol. 73, no. 2022, p. 101874, 2022, doi: <https://doi.org/10.1016/j.tele.2022.101874>.
- [16] P. I. Davidsen and F. Barnabè, “Exploring the potentials of behavioral system dynamics: insights from the field,” *J. Model. Manag.*, vol. 15, no. 1, pp. 339–364, Jan. 2019, doi: 10.1108/JM2-03-2019-0081.
- [17] F. D. Davis, “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology,” *MIS Q.*, vol. 13, no. 3, p. 319, 1989, doi: 10.2307/249008.
- [18] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, “User Acceptance of Information Technology: Toward a Unified View,” *Source MIS Q.*, vol. 27, no. 3, pp. 425–478, 2003, doi: 10.2307/30036540.
- [19] Y. T. Lin and T. C. Wang, “A Study of Primary Students’ Technology Acceptance and Flow State When Using a Technology-Enhanced Board Game in Mathematics Education,” *Educ. Sci.*, vol. 12, no. 11, 2022, doi: 10.3390/educsci12110764.
- [20] A. Dillon and M. G. Morris, “User Acceptance of Information Technology,” *Inf. Today*, vol. 31, no. January 2001, pp. 3–32, 1996, doi: 10.1006/imms.1993.1022.
- [21] R. Ibrahim, S. Masrom, R. C. M. Yusoff, N. M. . Zainuddin, and Z. I. Rizman, “Student acceptance of educational games in higher education,” *J. Fundam. Appl. Sci.*, vol. 9, no. 3S, pp. 809–829, 2018, doi: 10.4314/jfas.v9i3s.62.
- [22] K. Wan, V. King, and K. Chan, “Examining flow antecedents in game-based learning to promote self-regulated learning and acceptance,” *Electron. J. e-Learning*, vol. 19, no. 6, pp. 531–547, 2021, doi: 10.34190/ejel.19.6.2117.
- [23] M. Venter, “Continuance Use Intention of Primary School Learners Towards Mobile Mathematical Applications,” in *IEEE Frontiers in Education Conference (FIE)*, 2016. doi: 10.1109/FIE.2016.7757539.
- [24] K. Kiili, T. Lainema, S. De Freitas, and S. Arnab, “Flow framework for analyzing the quality of educational games,” *Entertain. Comput.*, vol. 5, no. 4, pp. 367–377, 2014, doi: 10.1016/j.entcom.2014.08.002.
- [25] M. Ashfaq, Q. Zhang, A. U. Zafar, M. Malik, and A. Waheed, “Understanding Ant Forest continuance: effects of user experience, personal attributes and motivational factors,” *Ind. Manag. Data Syst.*, vol. 122, no. 2, pp. 471–498, 2022, doi: 10.1108/IMDS-03-2021-0164.
- [26] R. Power, *Technology and the Curriculum: Summer 2019*. Lethbridge, AB, Canada: Power Learning Solutions, 2019. Available online: <https://techandcurr2019.pressbooks.com/> (accessed on 12 February 2023).
- [27] H. H. Razami, R. Ibrahim, and N. M. Aljuaid, “A Review of Models and Factors for Digital Educational Games Acceptance and User Experience,” *ISMSIT 2022 - 6th Int. Symp. Multidiscip. Stud. Innov. Technol. Proc.*, no. 21, pp. 37–42, 2022, doi: 10.1109/ISMSIT56059.2022.9932728.
- [28] A. H. Allam, A. Razak, and C. Hussin, “User Experience: Challenges and Opportunities,” *J. Res. Innov. Inf. Syst.*, pp. 28–36, 2009. Available online: http://seminar.spaceutm.edu.my/jisri/download/F_FinalPublished/Pub20_UserExperienceChallenges.pdf%5Cnhttp://seminar.utmspace.edu.my/jisri/download/F1_FinalPublished/Pub4_UserExperienceChallenges.pdf (accessed on 22 March 2023).
- [29] K. Poels, Y. A. W. de Kort, and W. A. Ijsselstein, “D3.3: Game Experience Questionnaire: Development of a self-report measure to assess the psychological impact of digital games,” *Tech. Univ. Eindhoven*, 2007. Available online: <https://www.semanticscholar.org/paper/D3.3-%3A-Game-Experience-Questionnaire%3Adevelopment-of-Poels-Kort/f9a9396fb48f6bcb942dc0a5cf7faaf7474dbe9> (accessed on 25 March 2023).
- [30] M. H. Phan and J. R. Keebler, “The Development and Validation of the Game User Experience Satisfaction Scale (GUESS),” *Hum. Factors Ergon. Soc.*, vol. 58, no. 8, pp. 1217–1247, 2016, doi: 10.1177/0018720816669646.
- [31] V. Nagalingam, R. Ibrahim, and R. Che Mohd Yusoff, “EDUGXQ: User Experience Instrument for Educational Games’ Evaluation,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 1, pp. 562–569, 2020, doi: 10.14569/IJACSA.2020.0110170.
- [32] R. F. DeVellis, *Scale development: Theory and applications*. SAGE Publications, Inc., 1991. Available online: <https://us.sagepub.com/en-us/nam/scale-development/book269114> (accessed on 23 January 2023).
- [33] F. R. López, M. Arias-Oliva, J. Pelegrín-Borondo, and L. M. Marín-Vinuesa, “Serious games in management education: An acceptance analysis,” *Int. J. Manag. Educ.*, vol. 19, no. 3, 2021, doi: 10.1016/j.ijme.2021.100517.
- [34] R. F. Malaquias, F. F. O. Malaquias, and Y. Hwang, “Understanding technology acceptance features in learning through a serious game,” *Comput. Human Behav.*, vol. 87, no. June, pp. 395–402, 2018, doi: 10.1016/j.chb.2018.06.008.
- [35] F. Fu, R. Su, and S. Yu, “EGameFlow: A scale to measure learners’ enjoyment of e-learning games,” *Comput. Educ.*, vol. 52, no. 152, pp. 101–112, 2009, doi: 10.1016/j.compedu.2008.07.004.
- [36] K. Cagiltay and S. Yeni, “A heuristic evaluation to support the instructional and enjoyment aspects of a math game,” *Progr. Electron. Libr. Inf. Syst.*, vol. 51, no. 4, pp. 406–423, Jan. 2017, doi: 10.1108/PROG-07-2016-0050.
- [37] R. D. Tobji, “Content Validity of Instruments in IS Research,” *J. Inf. Technol. Theory Appl.*, vol. 8, no. 3, pp. 31–56, 2006. Available online: <https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1057&context=jitta> (accessed on 7 February 2023).
- [38] D. F. Polit, C. T. Beck, and S. V. Owen, “Focus on Research Methods: Is the CVI an Acceptable Indicator of Content Validity? Appraisal and Recommendations,” *Res. Nurs. Health*, vol. 30, pp. 459–467, 2007, doi: 10.1002/NUR.20199.
- [39] N. Elangovan and E. Sundaravel, “Method of preparing a document for survey instrument validation by experts,” *MethodsX*, vol. 8, no. April, p. 101326, 2021, doi: 10.1016/j.mex.2021.101326.

- [40] C. H. Lawshe, "A Quantitative Approach To Content Validity," *Pers. Psychol.*, vol. 28, no. 4, pp. 563–575, 1975, doi: 10.1111/j.1744-6570.1975.tb01393.x.
- [41] A. Ibiyemi, Y. Mohd Adnan, M. N. Daud, A. Jogunola, and S. Olanrele, "A content validity study of the test of valuers' support for capturing sustainability in the valuation process in Nigeria," *Pacific Rim Prop. Res. J.*, vol. 25, no. 3, pp. 1–17, 2019, doi: 10.1080/14445921.2019.1703700.
- [42] J. Martí-Parreño, A. Galbis-Córdova, and M. J. Miquel-Romero, "Students' attitude towards the use of educational video games to develop competencies," *Comput. Human Behav.*, vol. 81, no. 2018, pp. 366–377, 2018, doi: 10.1016/j.chb.2017.12.017.
- [43] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, 3rd ed. Thousand Oaks: Sage, 2022. [Online]. Available: <https://www.pls-sem.net/pls-sem-books/a-primer-on-pls-sem-3rd-ed/>.
- [44] B. Luyt, Y. R. Guo, and D. H.-L. Goh, "Tertiary students' acceptance of a game to teach information literacy," *Aslib J. Inf. Manag.*, vol. 69, no. 1, pp. 46–63, Jan. 2017, doi: 10.1108/AJIM-08-2016-0131.
- [45] G. Sideris and S. Xinogalos, "Py-Rate Adventures: A 2D Platform Serious Game for Learning the Basic Concepts of Programming With Python," *Simul. Gaming*, vol. 50, no. 6, pp. 754–770, 2019, doi: 10.1177/1046878119872797.
- [46] Y. Tao, C. Cheng, and S. Sun, "What influences college students to continue using business simulation games? The Taiwan experience," *Comput. Educ.*, vol. 53, no. 3, pp. 929–939, 2009, doi: 10.1016/j.compedu.2009.05.009.
- [47] R. Ibrahim, S. Masrom, R. C. M. Yusoff, N. M. M. Zainuddin, Z. I. Rizman, and M. Sciences, "Student Acceptance of Educational Games in Higher Education," *J. Fundam. Appl. Sci.*, vol. 9, no. 3S, pp. 809–829, 2017, doi: <http://dx.doi.org/10.4314/jfas.v9i3s.62>.
- [48] Y. Wang, Y. Wang, and S. Jian, "Investigating the Determinants of Students' Intention to Use Business Simulation Games," *J. Educ. Comput. Res.*, vol. 58, no. 2, pp. 433–458, 2020, doi: 10.1177/0735633119865047.
- [49] R. F. Resende and C. I. P. S. Pádua, "Evaluating Software Engineering Simulation Games: the UGALCO framework," in *IEEE Frontiers in Education Conference (FIE) Proceedings*, 2014. doi: 10.1109/FIE.2014.7044204.
- [50] Y. M. Huang, "Students' Continuance Intention Toward Programming Games: Hedonic and Utilitarian Aspects," *Int. J. Hum. Comput. Interact.*, vol. 36, no. 4, pp. 393–402, 2020, doi: 10.1080/10447318.2019.1647665.
- [51] Y. Wirani, T. Nabarian, and M. S. Romadhon, "Evaluation of continued use on Kahoot! as a gamification-based learning platform from the perspective of Indonesia students," *Procedia Comput. Sci.*, vol. 197, no. 2021, pp. 545–556, 2022, doi: 10.1016/j.procs.2021.12.172.
- [52] C. Udeozor, F. Russo-Abegão, and J. Glassey, "Perceptions and factors affecting the adoption of digital games for engineering education: a mixed-method research," *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, 2023, doi: 10.1186/s41239-022-00369-z.

Human-object Behavior Analysis Based on Interaction Feature Generation Algorithm

Qing Ye¹, Xiuju Xu², Rui Li³

School of Information Science and Technology, North China University of Technology, Beijing, China^{1,2}
SNBC, Shandong, China³

Abstract—Aiming at the problem of insufficient utilization of interactive feature information between human and object, this paper proposes a two-stream human-object behavior analysis network based on interaction feature generation algorithm. The network extracts human-object's feature information and interactive feature information respectively. When extracting human-object features information, considering that ResNeXt has powerful feature expression ability, the network is used to extract human-object features from images. When extracting interactive features information between human and object, an interaction feature generation algorithm is proposed, which uses the feature reasoning ability of graph convolutional neural networks. A graph model is constructed by taking human and objects as nodes and the interaction between them as edges. According to the interactive feature generation algorithm, the graph model is updated by traversing nodes, and new interactive features are generated during this process. Finally, the humans' and objects' features information and the human-object interaction feature information are fused and sent to the classification network for behavior recognition, so as to fully utilize the humans' and objects' feature information and the interaction feature information of human-objects. The human-object behavior analysis network is experimentally verified. The results show that the accuracy of the network has been significantly improved on HICO-DET and V-COCO datasets.

Keywords—Two-stream human-object behavior analysis network; interaction feature generation algorithm; interactive feature information; ResNeXt; graph convolutional neural networks; graph model

I. INTRODUCTION

With the rapid development of science and technology, artificial intelligence (AI) has caused a profound impact in many fields [1], and its continuous improvement of related technologies [3] has penetrated into the application of various fields. In this dynamic context, research on human-object behavior analysis technology [5] has received increasing attention. How to effectively advance development of this technology will directly affect the application space of AI technology in real life. Whether it is possible to accurately and timely determine and analyze various interactions between people and objects in daily life will provide a solid and reliable foundation for further scene understanding and analysis, which will be of great value in theoretical research and engineering implementation.

However, due to the large amount of information, fast action and multiple interaction in the interaction process of human-objects, the specific interaction behaviors between

human and object cannot be accurately analyzed by utilizing all features, resulting in a low analysis rate of human-objects' behaviors.

In order to solve this problem, this paper researches on human-object behavior analysis, proposes an interactive feature generation network based on graph convolutional neural networks (GCNNs), and uses this network to analyze and identify the interaction between human and objects, trying to improve the accuracy of human-object behavior analysis on the basis of previous technologies.

The rest of this paper is organized as follows: Section II briefly reviews related work. Section III introduces the related models and algorithms of this paper. Section IV introduces the datasets used in the experiments, the experimental settings, and analyzes and discusses the experimental results. Finally, conclusions are drawn in Section V.

II. RELATED WORK

The traditional feature extraction method, which is to extract the behavioral features of human through the traditional manual method, has the advantages of simple implementation and strong operability. However, due to the relatively fixed templates used, it is difficult to perform fast and effective behavior analysis in facing complex environments and large datasets. Its application scenarios are limited, so more suitable for datasets with few types of behaviors and small scales. In order to improve the accuracy rate of human-object behavior analysis, it is often necessary to train on large datasets. Therefore, facing with such a huge amount of computation, graph convolution has certain advantages.

Simonyan et al. [7] proposed a feature extraction network algorithm based on two-stream convolutional neural networks (CNN). By extracting optical flow features from continuous images, and extracting image features from each frame, temporal and spatial features of continuous frame images are obtained, and two-channel feature information is fused and classified and identified. The proposed method breaks the concept of single-channel feature extraction, breaks through the limitations of feature extraction algorithms, and has a milestone significance. After that, based on multi-channel feature extraction, many researchers continued to improve and conduct in-depth research on human-object behavior analysis. However, these methods also increase a certain amount of computational burden in feature extraction.

Wang et al. [8] proposed a three-channel feature extraction network. In addition, some researchers proposed that human

skeleton points, as a means of posture prediction, can be used to supplement features. The multi-channel feature extraction algorithm formed by this method has gradually become research mainstream in the field of human-object behavior analysis. Wang et al. [9] continue to improve two-stream network architecture, and put forward advanced network architectures such as Temporal Segment Networks (TSN) for human-object behavior recognition.

He et al. [10] proposed Residual Network (ResNet) on previous Network research. This network proposes and optimizes the residual module, which not only deepens network depth, but also avoids training network degradation caused by network depth. The proposal of ResNet is of great significance in the field of image analysis. In the following years, researchers have continued to optimize and improve the ResNet, and proposed improved residual networks such as ResNeXt [26].

Yan et al. [11] and Mohamed et al. [12] have proposed spatial temporal graph convolutional networks (ST-GCN), which are used for general representation of skeleton sequences, realize human behavior analysis and recognition based on key points of human skeleton. This method constructs skeleton graph sequence by in space and time connecting the joint nodes of human skeleton, and builds multi-layer ST-GCN network based on this.

He et al. [13] proposed difficulty movement recognition method of calisthenics based on graph convolutional neural network (GCNN). This method constructs the multi-layer pyramid structure [14] of action images to complete the preprocessing of difficult movements in aerobics, and performs the training of samples to realize the difficult movement recognition of calisthenics. GCNN not only has more powerful expression ability and higher behavior recognition accuracy, but also has a stronger generalization ability, which makes it possible that the GCNN can be used for image feature extraction.

Thacker et al. [15] proposed through facial expression recognition to analyze human behavior. Aurangzeb et al. [16] proposed a human behavior analysis and recognition method based on multi-types features fusion and irrelevant features reduction, which initially selects a luminance channel and calculates motion estimation using optical flow. Afterwards, the moving regions are extracted through background subtraction approach. In the features extraction step, shape, color, and Gabor wavelet features are extracted and fused based on serial method. Thereafter, reduced irrelevant and redundant features are removed by Von Neuman entropy approach. The selected reduced features are finally recognized by One-Against-All (OAA) Multi-class SVM classifier.

Degardin et al. [17] provided a comprehensive overview of human behavior analysis over the past decade, presenting state-of-the-art and must-know methods in this field. Lanovaz [18] described machine behavior analysis. He proposed that machine behavior analysis is a science, which can study artificial behavior through its replicability, behavioral terms and philosophical assumptions of human behavior analysis, and study how machines interact with external environment and produce relevant changes. Lin et al. [19] built a framework

of external and internal factors to analyze human behavior, and game theory is used to simulate the conflicting interests of human behavior.

Bhatnagar et al. [20] proposed BEHAVE dataset, which is the first full body human-object interaction dataset with multi-view RGBD frames and corresponding 3D SMPL and object fits along with the annotated contacts between them. They also proposed a method that can record and track not just the humans and objects but also their interactions. Peng et al. [21] proposed a 3D max residual feature map convolution network (3D-MRCNN). The model can solve the deficiencies of the network degradation and gradient disappearance caused by convolution calculation, and achieve the improvement of classification accuracy without reducing the training efficiency, which is of value in the field of human behavior analysis in intelligent sensor networks.

In the field of human-object behavior analysis, there are various interactions between human and objects, the interaction between the same human and object may be different. For example, there are a variety of behavioral relations between human and bicycle, such as "riding", "pushing" and "resisting". In addition, it is also common to have multiple simultaneous interactions with the same human-object. For example, when a person picks up a glass to drink, he or she has both "lifting" and "drinking" interactive behaviors. In these cases, if interactive features between human-object aren't fully utilized, even if the human and object are identified, it is difficult to accurately and comprehensively analyze interactive behavior between them, resulting in inaccurate or omission analysis of human-object behavior. Obviously, if interaction features among human-objects aren't fully utilized, the accuracy rate of human-object analysis will be greatly reduced.

However, there are still some limitations in the use and reasoning of interactive behavior feature information in the existing methods for human behavior analysis. Therefore, in order to solve the above problems, this paper proposes a two-stream human-object behavior analysis network based on interactive feature generation algorithm. Although the method in this paper has made remarkable progress in the field of human-object behavior analysis, it still has some limitations in the behavior analysis of fuzzy or obscured human-objects, and its accuracy needs to be further improved.

III. METHOD

A. System Overview

In this paper, aiming at the problem of insufficient utilization of human-objects' interactive behavior features, a two-stream human-object behavior analysis network based on interaction feature generation algorithm is proposed. The network takes human-object features information obtained from the object detector as the input of the interaction feature generation network, and the interaction feature generation algorithm is used to generate new human-object behavior interaction features, thereby enhancing the full utilization of the interaction behavior features. The network block diagram is shown in Fig. 1. Drawing on the idea of "two-stream network" [7], a two-stream human-object behavior analysis network based on CNNs and GCNNs is constructed. This network is

divided into two streams to extract feature information of human and objects and the interaction feature information between the human-object from images. First of all, the input image is preprocessed. Then, the preprocessed image and the original image are fed into the object detector (e.g., Faster R-CNN), and the human-object features information in the image are extracted through the backbone of the object detector, such as ResNet. Human-object features information are input into the interaction feature generation network, the interactive features information of them are generated through this network. The specific method are as follows. Firstly, the human-object feature sequence is extracted by using CNN and input into the interactive feature generation network. Secondly, the detected people and objects are taken as nodes and the interactions between them are taken as edges to build an interactive feature graph model, which is used as the input of GCNN. Finally, the interactive feature generation algorithm is used to traverse the nodes of the graph model to update the interactive relationship, and then generate new interactive features of human-object behaviors, so as to make full use of interactive behavioral features. Moreover, the humans' and objects' features information and the human-object interaction feature information are fused and sent to SoftMax classifier [22] for identification and classification, so the recognition result is obtained, which realizes the full use of human-object feature information and human-object interaction feature information.

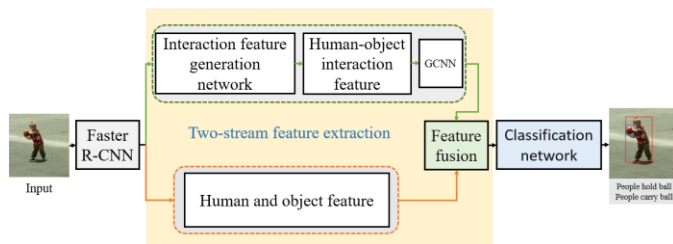


Fig. 1. Overall framework of model.

B. Image Preprocessing

In this paper, image preprocessing involves applying self-transformations to the images, generating new images that are different from the original ones, while preserving the original image category labels. Fig. 2 shows some examples of image preprocessing methods. That is, the image dataset can be augmented by operations such as horizontal flipping, vertical flipping, scaling, random rotation, and generative adversarial networks (GANs) [23] to expand the data samples.



Fig. 2. Example of image preprocessing method.

To reduce computational complexity, increase the training data samples, and improve the model's generalization ability, the original input image is preprocessed before object detection. Specifically, the original image is first input. Next, apply random rotation and flip to generate a new image by randomly rotating and flipping the original input image vertically, horizontally, or both. Then, the image is randomly cropped. Finally, adjust the height and width of all images to 400 x 400. In order to ensure the accuracy and adequacy of the original image and annotation data, no other processing or enhancement operations are applied to the images in this study.

C. Object Detection Network

This study adopts a two-stage approach to detect and analyze human-object behavior, which offers the advantage of reducing irrelevant background and interference from other human, while accurately utilizing and extracting features of the target human and object. Specifically, an object detector is first used for preliminary screening, and regions of interest for human and objects in image are proposed, thereby extracting effective bounding boxes for human and objects. Then, behavior analysis and inference are conducted on the selected humans and objects. To quickly and accurately locate and identify humans and objects in the image, this paper adopts Faster R-CNN [24] as the object detector.

Faster R-CNN is proposed based on the improved Fast R-CNN algorithm [25], which uses Region Proposal Networks (RPN) to replace original Selective Search (SS) method to generate proposal box, and CNN that generates proposal box and object detection are shared, which effectively improves the speed of detection and achieves good results. The whole Faster R-CNN system is a single and unified network, and RPN module is its "attention".

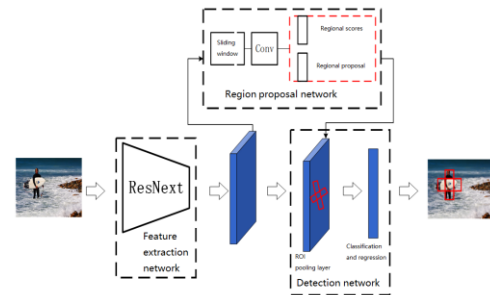


Fig. 3. Network framework of Faster R-CNN.

The network framework of Faster R-CNN is shown in Fig. 3. First, the pre-processed image is input into the feature extraction network (e.g., ResNeXt) to extract the human-object feature information and obtain the image feature map, which is shared by the RPN and the detection network. Second, the extracted feature map is input into RPN, and the sliding window is used to perform convolution operation with the input feature map. According to different scales and different basic sizes, a point on the feature map corresponds to k proposals on the original image. Classification judgment and regression operation are performed on the generated boxes respectively, and $2k$ regional scores and $4k$ regional proposals are obtained. After screening, a relatively accurate region candidate box is finally obtained. Finally, the feature map

output by the feature extraction network and the region candidate box information output by the region proposal network are integrated. And a fixed-size proposal feature map is generated through the RoI pooling layer, and that is input into the fully connected layer. The classification and regression network are used for classification and regression training to obtain the accurate position of the predicted object, so as to realize human-objects' identification and detection. The following sections will analyze each part of the Faster R-CNN in detail.

1) *Feature extraction network:* ResNext [26] is selected as the feature extraction network of Faster R-CNN. Its input is a pre-processed image x , and its output is a convolution feature map containing global features. Its main algorithm is on convolution operations. Convolution is a linear operation, and the specific operation process is shown in Formula (1):

$$s(t) = (x * w)(t) \quad (1)$$

Where x is the input of convolution layer, w is the kernel function, t is the time, and $s(t)$ is the output of convolution layer. When t is set to discrete time, the formula can be expressed as Formula (2):

$$s(t) = \sum_{-\infty}^{\infty} x(a)w(t - a) \quad (2)$$

a represents the discrete time point, $x(a)$ represents the picture x obtained at time a . In this paper, we need to perform convolution operation on image data. The computer sees the input image as a collection of pixels and transforms it into a two-dimensional digital matrix. When the convolution object is two-dimensional, the convolution kernel is also two-dimensional. The process of the convolution operation on the image is expressed by Formula (3):

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n) \quad (3)$$

In this paper, ResNeXt is selected as the backbone network of the object detector, whose network topology is shown in Fig. 4(a). ResNet [10] introduces a residual network topology to alleviate the gradient explosion problem during deep network training. However, as a deep network, although it can better extract features, the increase of network parameters will inevitably bring heavy computational burden, as shown in Fig.4(b). ResNeXt improves the topology of ResNet. It refers to the idea of replacing large convolution kernels with several small convolution kernels in VGG [27] and the idea of splitting in GoogLeNet. It adopts the algorithm of grouping convolution, which realizes that number of network layers and parameters are basically unchanged compared with ResNet, but the recognition accuracy is significantly improved.

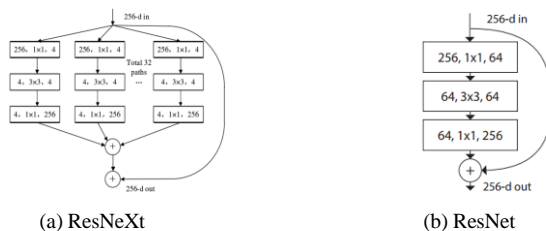


Fig. 4. Network topology.

The channel in Fig. 4(a) actually changes from 256 to 128 and then to 256, and the number of parameters is 70,144. In Fig. 4(b), the number of ResNet channels changes from 256 to 64 and then back to 256, and the number of parameters is 69,632. However, if the ResNet channel is also changed from 256 to 128 and then to 256, the number of parameters is 212,992, which is much larger than the above two values. Therefore, ResNeXt deepens the depth of feature maps in the convolution layers by grouping convolutions without increasing the number of parameters. ResNeXt combines advantages of GoogLeNet and residual network to improve accuracy without changing model complexity. ResNeXt is easy to train and can effectively avoid exploding gradient problem caused by network deepening. So, it is suitable for feature extraction.

2) *Region proposal network:* Region Proposal Network (RPN) is a fully convolutional network, which can demarcate human and objects in images and generate bounding boxes. RPN introduces the anchor mechanism, which uses a sliding window to map the anchor points of the feature map to the original image, and each anchor point is mapped to generate k region boxes with different scales and sizes. Intersection over union (IoU) between each region box and the desired region is calculated. When the IoU is greater than 0.7, it is marked as a positive sample; when the IoU is less than 0.3, it is marked as a negative sample. Regression training is performed on these samples, and then the calibrated region proposal boxes are screened and modified. RPN performs a binary classification judgment on whether there are human-objects in the generated k region boxes. If the result is that there are human-objects, the position of the region proposal box is modified according to the regression result. After iterative training, a number of bounding boxes are generated, and the non-maximum suppression algorithm is used to remove the overlapping bounding boxes to get the final effective human and object bounding boxes in images.

3) *Detection network:* The detection network consists of RoI pooling layer and classification regression network. The main function of the RoI pooling layer is to unify the corresponding features of the bounding boxes through pooling. The feature generated by the ROI pooling layer with a dimension of $c \times 7 \times 7$ (c is the number of channels, generally the number of object classes in the dataset) is reshaped into a vector. Send the vector to two full connected layers, and then conduct Softmax to obtain the probability of the object for different classes.

D. Interaction Feature Generation Algorithm

In traditional deep learning, data samples are often considered to be independent. But in graph neural networks (GNNs), each sample node will establish a connection with other data samples through edges, and this connection can be used to form interdependence between different instances [28]. So GNNs has powerful reasoning ability and feature propagation ability.

In this paper, to solve the problem of insufficient use of interactive information between human and objects in the images, an interactive feature generation algorithm based on GCN is proposed. Based on this algorithm, an interactive feature generation network is constructed, and the network framework is shown in Fig. 5. First of all, the pre-processed image is sent to the object detector (e.g., Faster R-CNN), which is used to conduct preliminary detection and segmentation of the image, and obtain the feature sequences of human and object respectively. Secondly, the graph model is constructed by taking people and objects as nodes and the interaction between them as edges, and the graph model is input into the interactive feature generation network in the form of data structure. Finally, each node in the graph model is traversed by the interactive feature generation algorithm, its interaction relationship is updated, and then new interactive features are generated.

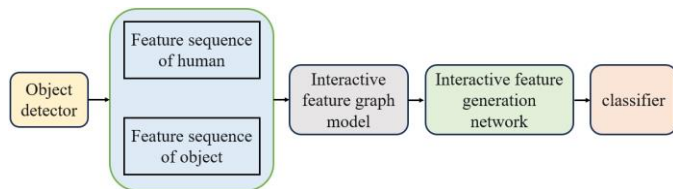


Fig. 5. Interaction feature generation network.

A graph model is a data structure composed of nodes and edges. Nodes refer to human and objects instances in the image, and edges refer to the relationship between them. Each node has its own features and structure information. The graph model is generally represented by an adjacency matrix, and the process is shown in Fig. 6. The calculation Formula (4) of the graph convolution operator is as follows, and the central node is set as i :

$$h_i^{l+1} = \sigma \left(\sum_{j \in N_j} \frac{1}{C_{ij}} h_j^l W_{R_j}^l \right) \quad (4)$$

Where h_i^{l+1} represents the feature representation of node i at the $l + 1$ layer; C_{ij} represents normalized factors, such as taking the reciprocal of the node degree; N_j represents the neighbor of node j , and contains its own information; R_j represents the type of node j ; $W_{R_j}^l$ represents the transformation weight parameter of a node of type R_j .

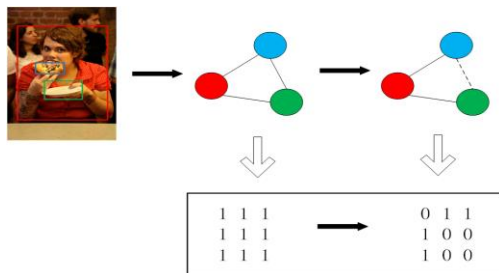


Fig. 6. Construction of graph model.

The construction process of the graph model is shown in Fig. 6. Specifically, humans and objects in the image are taken as nodes, and the interactions between them are taken as edges. Based on the interaction between human and objects, a fully

connected undirected graph G is constructed according to Formula (5), and undirected graph is initialized as an adjacency matrix with all elements being 1. Then, each node in the undirected graph is traversed through the interaction feature generation network. During the traversal process, according to whether there is the interaction between human and objects, the undirected graph and the corresponding adjacency matrix are updated.

$$G = (V, E) \quad (5)$$

Where V represents all nodes in the image, E represents all edges. Each object represents a node $v \in V$ of the undirected graph, and the interactive between different target objects is regarded as an edge $e \in E$. If there are n objects in the image, the undirected graph has $n(n - 1)/2$ edges. Since the graph model is converted into the form of adjacency matrix for calculation operation, the undirected graph is represented as an adjacency matrix $A_{n \times n}$ of $n \times n$ size, whose matrix elements $i, j \in \{0, 1\}$, 0 indicates that there is no interaction between node i and j , and 1 indicates that there is interaction.

Aiming at the problem of underutilization of human-object interaction behavior features, a two-stream human-object interaction analysis network based on interactive feature generation algorithm is proposed. A network example is shown in Fig. 7. Specifically, assume that there is an interaction between the person in the image and all objects in the image. However, in reality, there is generally no interaction between human and objects without overlap in space. Hence, in the interactive feature network, graph model and the corresponding adjacency matrix are updated according to the principle that there is no interaction between people and objects with non-overlapping bounding boxes in space. The updated graph model contains only the possible interactions between human and objects in the image. This not only saves computation, but also facilitates the subsequent graph reasoning.

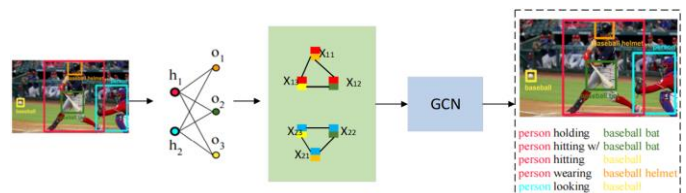


Fig. 7. Example of interaction feature generation algorithm.

The interaction feature generation algorithm learns the structural connection between human and objects, which is shown in Fig. 7. In the process of traversing nodes, better interactive features are generated according to the interaction between human and object, and the update process is shown in Formula (6):

$$x_{ho}^{(n+1)} = \sigma(x_{ho}^{(n)} + \sum_{o' \in O} W x_{ho'}^{(n)}) \quad (6)$$

Where h represents object judged to be human; o represents object judged as object, O represents the total number of objects; W is the projection matrix; $x_{ho}^{(n)}$ is the interaction relation between h and o , $x_{ho'}^{(n)}$ is the interaction between h and o' ($o' \in O$ and $o' \neq o$), $x_{ho}^{(n+1)}$ is the updated interaction between h and o .

The specific interaction feature generation algorithm formula is as follows:

$$F_h = f_h + \sum_{o=1}^O X_{ho} W_{oh}(f_o) \quad (7)$$

$$F_o = f_o + \sum_{h=1}^H X_{oh} W_{ho}(f_h) \quad (8)$$

As shown in Formulas (7) and (8), F represents the newly generated interactive feature vector, f represents the feature vector extracted from the original image, H represents total number of humans in image, and X represents the interaction between objects and human, and $X_{ho} = X_{oh} = x_{ho}^{(n+1)}$.

The generated human-object interaction features and human and object features generated by the object detector are sent to MLP for feature fusion. The fused features are fed into SoftMax classifier to obtain the final human-object behavior analysis result.

IV. EXPERIMENT

In this section, the experimental datasets and implementation details are first described. Then, our model is evaluated by quantitative comparison with state-of-the-art methods, followed by ablation studies to validate the components in our framework. Finally, several visualization results are shown to demonstrate effectiveness of our method.

A. Datasets

1) *HICO-DET* [29]: The HICO-DET dataset, introduced in 2018, is a large dataset for studying human-object behavior. It consists of 117 common behaviors involving 80 different objects, along with their associated behaviors. The sample are shown in Fig. 8 (a). The dataset contains a total of 47,776 images, with 38,118 images allocated to the training set and 9,658 images in the test set. A noteworthy aspect of this dataset is that many images feature multiple pairs of human-object behavior annotations, resulting in over 150,000 human-object pairs and 600 distinct human-object behavior categories in total.

2) *V-COCO* [30]: V-COCO is a specialized dataset in the field of human-object behavior analysis, which is derived from MS-COCO. The dataset is divided into three main parts: the training set comprising 2,533 images, the validation set containing 2,867 images, and the test set with 4,946 images. It not only inherits all the annotations from the MS-COCO dataset, but also incorporates additional semantic extension markup of human-object pairs through the use of AMT (Amazon Mechanical Turk) proposed by Gupta et al. This extension results in a more professional and refined sub-dataset. The sample are shown in Fig. 8 (b).

B. Implementation Details and Evaluation Metric

In this paper, Faster R-CNN [24] is used as object detector, its backbone module uses ResNeXt50 trained in ImageNet-1 K dataset as feature extractor. The SGD optimizer is used for training with an initial learning rate of 0.01, weight decay of 0.0001, and momentum of 0.9. For V-COCO, the learning rate is reduced to 0.001 at 80 iterations. For HICO-DET, the learning rate is reduced to 0.001 at 60 iterations. All experiments in this paper are conducted on GeForce RTX 2080Ti GPU and CUDA 11.4 with a batch size of 4.

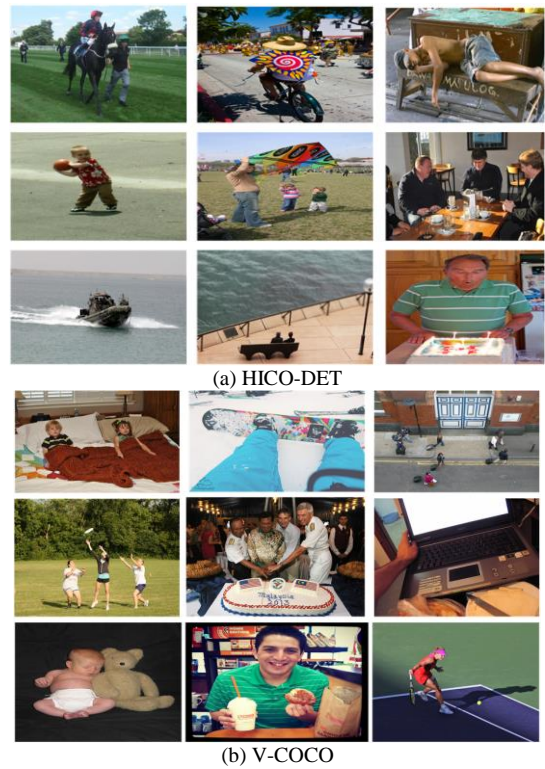


Fig. 8. Example images of datasets.

This paper uses accuracy to measure the performance of human-object behavior analysis. During accuracy calculation of the metrics, the prediction of the human-object pair is considered correct (1) if the IoU of the human-object bounding box and the ground-truth box is greater than 0.5, and (2) the interaction class label of the prediction of the human-object pair is correct.

C. Experimental Results

To analyze the effectiveness of our method in human-object behavior analysis in more detail, our proposed framework was compared with several existing human-object interaction detection methods on the V-COCO and HICO-DET datasets, and the results are shown in Table I.

TABLE I. COMPARISON OF ACCURACY RATE OF HUMAN-OBJECT BEHAVIOR ANALYSIS METHODS

Human-object behavior analysis methods	V-COCO (%)	HICO-DET (%)
Wang[31]	47.3	21.08
DRG[32]	51.0	23.89
TIN[33]	47.8	20.26
PMFNet[34]	52.0	21.20
PFNet[35]	52.8	24.89
PPDM[36]	-	26.84
IP-Net[37]	51.0	23.92
UnionDet[38]	47.5	21.27
Ours method	52.4	27.89

By comparison and analysis of the data in Table I, it can be seen that in recent years, the accuracy of most human-object behavior analysis methods on the V-COCO database has been generally more than 50%, among which the PFNet method has the highest accuracy of 52.8%, which is 0.4% higher than our method. One important reason is that PFNet uses more fine-grained body part level information. However, because this paper uses a graph model structure with strong inference ability, and design a graph node adaptive interactive update algorithm, so the results are not very different. Although the accuracy of our method is lower than that of PFNet method, it has a significant increase compared with other methods.

According to the data in Table I, the accuracy of human-object behavior analysis on the HICO-DET database is generally low. However, our method achieves the state-of-the-art results on the HICO-DET database, which is 1.05% higher than the previous state-of-the-art the PPDM method. It is inferred that the reason may be the difference in the difficulty and the focus of dataset. Compared with the existing human-object behavior analysis, the accuracy rate of the two-stream network has been significantly improved.

D. Ablation Study

In this paper, a two-stream human-object behavior analysis network based on interaction feature generation algorithm is proposed and its effectiveness is tested. The first experiment is to test the feature extraction networks in object detector. ResNeXt 50 and ResNet 50 are used as the feature extraction networks separately. Training and testing are carried out on the HICO-DET and the V-COCO datasets. The suitable extraction network is selected by comparing the test results on each dataset. The second experiment is to verify the effectiveness of network proposed in this paper. In this experiment, three networks are proposed: (1) a single-stream human-object behavior analysis network without using interaction feature generation algorithm; (2) a single-stream human-object behavior analysis network with interaction feature generation algorithm; (3) a two-stream human-object behavior analysis network based on interaction feature generation algorithm. Training and testing are carried out on the HICO-DET and V-COCO datasets respectively, and through the test results on each dataset, it is proved whether the interaction feature generation algorithm and the two-stream human-object behavior analysis network based on the interaction feature generation algorithm are beneficial to improve the accuracy rate of human-object behavior analysis.

1) *Comparison experiment between ResNeXt 50 and ResNet 50:* In this experiment, ResNeXt 50 and ResNet 50 are used as the backbone of the object detector, respectively, training and testing are carried out in two databases. ResNeXt 50 and ResNet 50 pre-trained models are tested in the new database by means of transfer learning. Fig. 9 and Fig. 10 show the experiments performed on V-COCO and HICO-DET databases, respectively.

As shown in Fig. 9, the horizontal axis represents the number of training epochs, and the vertical axis represents the accuracy of the test set for 29 kinds of human-object behavior analysis in the V-COCO database. The green and red curves

represent the accuracy test results of ResNeXt 50 and ResNet 50 as the number of training epochs increases, respectively.

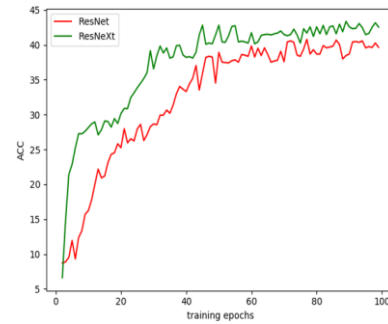


Fig. 9. Comparison of accuracy of classification results in V-COCO database.

As shown in Fig. 10, the horizontal axis represents the number of training epochs, the vertical axis represents the test sets accuracy of 117 kinds of human-object behavior analysis in HICO-DET database. The green and red curves represent the accuracy test results of ResNeXt 50 and ResNet 50 as the number of training epochs increases, respectively.

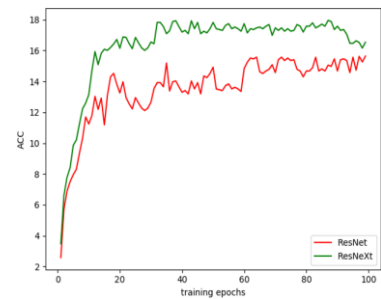


Fig. 10. Comparison of accuracy of analysis results in HICO-DET database.

It can be seen from the experimental results in Fig. 9 and Fig. 10, the number of training epochs of the network model is the same, and the accuracy rate of the human-object behavior analysis result of ResNeXt 50 is better than that of ResNet 50. In HICO-DET database, when ResNeXt 50 is used as the backbone of the object detector, the convergence is better, and its accuracy is significantly higher than that ResNet 50. The use of ResNeXt 50 as backbone for object detector in human-object behavior analysis can be concluded to significantly improve accuracy.

2) *Verify the effectiveness of the two-stream human-object behavior analysis network based on interaction feature generation algorithm:* Based on the above experimental analysis, this study decided to use ResNeXt 50 as the feature extraction network of object detector. Therefore, in the second experiment, in order to verify the effectiveness of the two-stream network, training and accuracy test are carried out on the single-stream network without using the interactive feature generation algorithm, the single-stream network using interactive feature generation algorithm, and the two-stream network based on the interactive feature generation algorithm on the two databases respectively.

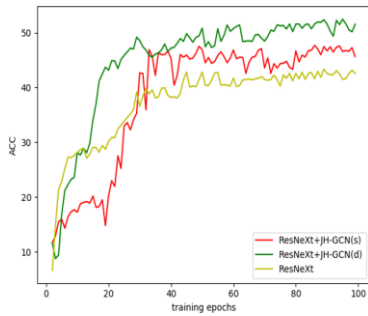


Fig. 11. Effectiveness comparison experiment in V-COCO database.

As shown in Fig. 11, the three proposed network structures are trained and tested in the V-COCO database. The horizontal axis represents training epochs, and the vertical axis represents the accuracy rate of the 29 human-object behavior test sets in the V-COCO database. The green, red and yellow curves respectively represent the training and accuracy testing process on the two-stream network, the single-stream network using the interactive feature generation algorithm, and single-stream network without using interactive feature generation algorithm.

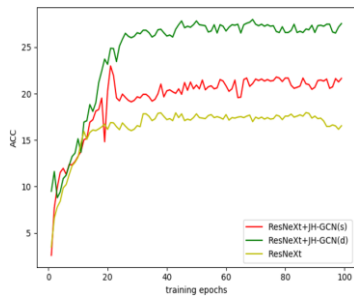


Fig. 12. Effectiveness comparison experiment in HICO-DET database.

As shown in Fig. 12, the horizontal axis represents training epochs, the vertical axis represents accuracy of the analysis of 117 common behaviors in the HICO-DET database. The green, red and yellow curves respectively represent the training and accuracy testing process on the two-stream network, the single-stream network using interaction feature generation algorithm, and single-stream network without using interaction feature generation algorithm.

The experimental results of training and testing on two datasets are analyzed in Fig. 11 and Fig. 12. The overall trend of the curve is as follows. The green curve is above the red and yellow curve, and the red curve is above the yellow. In the V-COCO database, there are coincident points before the curves converge. But after convergence, the convergence accuracy of the two-stream network is higher than the other two networks, and the convergence speed is faster. The accuracy rate of the single-stream network using the interaction feature generation algorithm is higher than that of the single-stream network without using interaction feature generation algorithm. In the HICO-DET database, all three network curves converge quickly. It is obvious that the accuracy of the two-stream network after convergence is higher than that of other two networks. And the accuracy of the single-stream network with interaction feature generation algorithm is higher than that of the single-stream network without using the interaction feature generation algorithm.

From the above analysis, it can be concluded that the interaction feature generation algorithm can effectively improve the accuracy of human-object behavior analysis, and accuracy of the two-stream network is higher than that of other two networks, which indicates the effectiveness of the two-stream network in human-object behavior analysis.

In this paper, four network architectures are trained and tested in the V-COCO and HICO-DET databases. The single-stream network without the interaction feature generation algorithm and the feature extraction network is ResNet 50, the single-stream network without interaction feature generation algorithm and the feature extraction network is ResNeXt 50, the single-stream network with interaction feature generation algorithm and the extraction network is ResNeXt 50, and the two-stream network based on interaction feature generation algorithm (the feature extraction network is ResNeXt 50).

TABLE II. COMPARISON OF ACCURACY OF FOUR NETWORKS

The network architecture	V-COCO	HICO-DET
ResNet 50	40.3	15.73
ResNeXt 50	43.2	17.71
ResNeXt 50+ Interaction feature generation algorithm	47.6	21.54
A two-stream human-object behavior analysis network based on interaction feature generation algorithm	52.4	27.89

The specific accuracy the human-object behavior analysis methods of the four network architectures are shown in Table II. When ResNet 50 is used as the feature extraction network of the object detector, the model obtained after training on the V-COCO database is used to test, its accuracy is 40.3%; the model obtained after training on the HICO-DET database is used to test, its accuracy is 15.73%. When ResNet 50 is used as the feature extraction network of the object detector, the model obtained after training on the V-COCO database is used to test, its accuracy is 43.2%, which is 2.9% higher than ResNet 50; the model obtained after training on the HICO-DET database is used to test, its accuracy is 17.71%, which is 1.98% higher than ResNet 50. It can be concluded that ResNeXt 50 is more suitable as the feature extraction network of the object detector.

On the basis of determining that the feature extraction network is ResNeXt 50, according to the proposed interaction feature generation algorithm, the single-stream and the two-stream human-object behavior analysis network based on the interaction feature generation algorithm are proposed. The test results are shown in Table II. For V-COCO database, the model obtained after training of the single-stream network using the interaction feature generation algorithm is tested, its accuracy rate is 47.6%, which is 4.4% higher than that without the interaction feature generation algorithm. The model obtained after training on the HICO-DET database is used to test, its accuracy is 21.54%, which is 3.83% higher than that without the interaction feature generation algorithm. The experimental data show that interaction feature generation algorithm is beneficial to improve the accuracy of human-object behavior analysis.

For V-COCO database, the model obtained after training of the two-stream network is tested, its accuracy is 52.4%, which is 4.8% higher than the single-stream network using interaction feature generation algorithm. The model obtained after training on HICO-DET database is used to test, its accuracy is 27.89%, which is 6.35% higher than that the single-stream network using the interaction feature generation algorithm. As shown in Table II, the accuracy rate of the two-stream network based on interaction feature generation algorithm is higher than that of the other three networks, which shows that the network is effective in human-object behavior analysis.

E. Visualized Results

To qualitatively visualize the detection effect of our model, Fig. 13 shows the visualization results of human-object behavior analysis on the test images of our model.



Fig. 13. Example detections of human-object behavior analysis method based on interaction feature generation algorithm.

In the nine example detections images in Fig. 13, after the human and objects are detected in each image, the behavior analysis results are successively, "catch", "ride", "ride", "drink", "ski" and "look", "talk_on_phone", "kick", "eat_obj", "brush". Human and objects as background aren't selected. It can be concluded that the human-object behavior analysis method based on the interaction feature generation algorithm proposed in this paper is effective.

As can be seen from the above experiments, the experimental results validate the effectiveness of our proposed model in enhancing human-object behavior analysis performance. The comparison ablation study and qualitative results collectively demonstrate the superior performance and robustness of our proposed method on V-COCO and HICO-DET datasets. Therefore, the algorithm can be used in areas such as intelligent surveillance video and intelligent robots.

V. CONCLUSION

Aiming at the problem of insufficient utilization of interactive behavior feature information between human and object, a two-stream human-object behavior analysis network based on interaction feature generation algorithm is proposed. The network uses interactive feature generation algorithm to generate new behavior interaction features, so as to make the full use of interactive behavior features. After experimental verification on datasets, the recognition accuracy of the two-stream network based on interaction feature generation algorithm is higher, compared with the single-stream network

that only uses the image feature extraction network and only uses the interaction feature generation algorithm.

However, the proposed algorithm still has some limitations when it comes to the behavior analysis of fuzzy or obscured human-objects in images. Therefore, before conducting human behavior analysis, additional techniques are required to accurately process and identify blurred or obscured people or objects. Future research efforts will focus on improving the processing of difficult and fuzzy human-object recognition processes, aiming to optimize human-object behavior analysis techniques and improve the accuracy of human-object behavior analysis.

ACKNOWLEDGMENT

This paper is supported by Key Research and Development Program (No.2020YFC0811004), Key Research and Development Program (No.2020YFB1600702), Technology Project of Beijing Municipal Education Commission (No. SQKM201810009002), Beijing Innovation Team, Key scientific research direction construction project of North China University of Technology. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

- [1] Y. Xu, The Vigorous Development of Artificial Intelligence Innovation Application[N]. People's Posts and Telecommunications News, 2022-05-10(001).
- [2] B. Zhang, J. Zhu, H. Su, "Toward the third generation artificial intelligence." Science China Information Sciences 66.2 (2023): 121101.
- [3] X. Zhang, "Application of artificial intelligence recognition technology in digital image processing." Wireless Communications and Mobile Computing 2022 (2022): 1-10.
- [4] J.H. Tao, J.T. Gong, N. Gao, S.W. Fu, S. Liang, C. Yu, Human-computer interaction oriented to virtual-real integration. Chinese Journal of Image and Graphics, 2023, 28(06):1513-1542.
- [5] X. Yun, H.S. Song, H.X. Liang, et al., Behavior Analysis System for Key Positions Based on Deep Learning[J]. Computer Engineering and Applications, 2021, 57(06): 225-231.
- [6] M.L. Deng, Z.D. Gao, L. Li, et al., A review of human behavior recognition based on deep learning[J]. Computer Engineering and Applications, 2022, 58(13): 14-26.
- [7] K. Simonyan, A. Zisserman, Two-stream convolutional networks for action recognition in videos[J]. Advances in neural information processing systems, 2014, 1(4): 568-576.
- [8] H. Wang, C. Schmid, Action recognition with improved trajectories[C]. Proceedings of the IEEE international conference on computer vision, 2013: 3551-3558.
- [9] L. Wang, Y. Xiong, Z. Wang, et al., Temporal segment networks for action recognition in videos[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(11): 2740-2755.
- [10] K. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778..
- [11] S. Yan, Y. Xiong, D. Lin, Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. 32nd AAAI Conference on Artificial Intelligence, 2018: 7444-7452.
- [12] A. Mohamed, K. Qian, M. Elhoseiny, et al., Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14424-14432.
- [13] L. He, H.M. Li, J.G. Sun, et al., Difficulty movement recognition method of calisthenics based on graph convolutional neural network[J]. Journal of West Anhui University, 2022, 38(02): 136-141.

- [14] S.M. Pan, Y.J. Wang, Y.W. Zhong, Cross-domain pedestrian re-identification based on graph convolutional neural network[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2020, 48(09): 44-49.
- [15] C.B. Thacker, R.M. Makwana, Human behavior analysis through facial expression recognition in images using deep learning[J]. International Journal of Innovative Technology and Exploring Engineering, 2019, 9(2): 391-397.
- [16] K. Aurangzeb, I. Haider, M.A. Khan, et al., "Human behavior analysis based on multi-types features fusion and Von Nauman entropy based features reduction." Journal of Medical Imaging and Health Informatics 9.4 (2019): 662-669.
- [17] B. Degardin, H. Proença, Human behavior analysis: a survey on action recognition[J]. Applied Sciences, 2021, 11(18): 8324.
- [18] M.J. Lanovaz, Some Characteristics and Arguments in Favor of a Science of Machine Behavior Analysis[J]. Perspectives on behavior science, 2022, 45(2): 399-419.
- [19] A. Lin, Y. Xu, H. Shen, Quantitative Analysis of Human Behavior in Environmental Protection[J]. Journal of the Knowledge Economy, 2022: 1-28.
- [20] B.L. Bhatnagar, X. Xie, I. A. Petrov, C. Sminchisescu, C. Theobalt, and G. Pons-Moll, (2022). Behave: Dataset and method for tracking human object interactions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15935-15946.
- [21] B. Peng, Z. Yao, Q. Wu, H. Sun and G. Zhou, (2022). 3D convolutional neural network for human behavior analysis in intelligent sensor network. Mobile Networks and Applications, 27(4), 1559-1568.
- [22] G.X. Liu, J. Huang, Transfer learning technology of machine vision detection discriminative semantic segmentation based on label reservation Softmax algorithm[J]. Optical Precision Instruments, 2022, 30(01): 117-125.
- [23] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.
- [24] H. Duan, J. Huang, W. Liu and F. Shu, (2022, August). Defective Surface Detection based on Improved Faster R-CNN. In 2022 IEEE International Conference on Industrial Technology (ICIT), pp. 1-6. IEEE.
- [25] R. Girshick, Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [26] Z. Zhang, Research on face recognition method and application based on single sample[D]. University of Electronic Science and Technology of China, 2021.
- [27] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409-1556, 2014.
- [28] X.R. Wang, H. Zhang, Small sample classification network based on attention mechanism and graph convolution[J]. Computer Engineering and Applications, 2021, 57(19): 164-170.
- [29] Y.W. Chao, Y. Liu, X. Liu, H. Zeng and J. Deng, (2018, March). Learning to detect human-object interactions. In 2018 IEEE winter conference on applications of computer vision (wacv), pp. 381-389. IEEE.
- [30] S. Gupta, J. Malik, Visual semantic role labeling[J]. arXiv preprint arXiv:1505-04474, 2015.
- [31] H. Wang, W.S. Zheng and Y.B. Ling, "Contextual heterogeneous graph network for human-object interaction detection." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16. Springer International Publishing, 2020.
- [32] C. Gao, J. Xu, Y. Zou and J.B. Huang, "Drg: Dual relation graph for human-object interaction detection." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16. Springer International Publishing, 2020.
- [33] Y. L. Li, S. Zhou, X. Huang, L. Xu, Z. Ma, H. S. Fang, et al., (2019). Transferable interactiveness knowledge for human-object interaction detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3585-3594.
- [34] B. Wan, D. Zhou, Y. Liu, R. Li and X. He, (2019). Pose-aware multi-level feature network for human object interaction detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9469-9478.
- [35] H. Liu, T. J. Mu and X. Huang, (2021). Detecting human-object interaction with multi-level pairwise feature network. Computational Visual Media, 7, 229-239.
- [36] Y. Liao, S. Liu, F. Wang, Y. Chen, C. Qian and J. Feng, (2020). PPDm: Parallel point detection and matching for real-time human-object interaction detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 482-490.
- [37] T. Wang, T. Yang, M. Danelljan, F. S. Khan, X. Zhang, J. Sun, (2020). Learning human-object interaction detection using interaction points. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4116-4125.
- [38] B. Kim, T. Choi, J. Kang and H.J. Kim, (2020). Uniondet: Union-level detector towards real-time human-object interaction detection. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16 (pp. 498-514). Springer International Publishing.

A Proposed Framework for Context-Aware Semantic Service Provisioning

Wael Haider¹, Hatem Abdelkader², Amira Abdelwahab³

Department of Management Information Systems, Higher Institute of Qualitative Studies, Heliopolis, Cairo 11757, Egypt¹

Department of Information Systems-College of Computers and Information,
Menoufia University, Shibin Al Kawm 32511, Menoufia, Egypt^{1,2,3}

Department of Information Systems-College of Computer Sciences and Information Technology,
King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia³

Abstract—Web-hosted Internet of Things (IoT) applications are the next logical step in the recent endeavor by academia and industry to design and standardize new communication protocols for smart objects. Context Awareness is defined as the property of a system that employs context to provide related information or services to the user, where the relationship is based on the user's task. Therefore, context-aware service discovery can be defined as utilizing context information to discover the most relevant services for the user. Merging context-aware concepts with the IoT facilitates IoT system developments that depend on complex environments with many sensors and actuators, user, and their environment. The main objective of this study is to design an abstract framework for provisioning smart objects as a service based on context-aware concepts while considering constraints of bandwidth, scalability, and performance. The proposed framework building blocks include data acquisition and management service and data aggregation, and rules reasoning. The proposed framework is validated and evaluated by constructing an IoT network simulation and testing accessing the service in the traditional method and according to the proposed framework and comparing the results.

Keywords—Internet of Things (IoT); Web of Things (WoT); Web of Objects (WoOs); context-awareness; service provisioning; interoperability; ontology; OWL

I. INTRODUCTION

Every object in our environment, including chairs, gas meters, electricity meters, curtains, lights, office equipment, and home appliances, should be transformed into Internet-connected smart objects to improve a variety of application domains (e.g., building automation, healthcare services, smart grids, transportation, and environmental monitoring).

A smart object is defined as an entity that is provided with a sensor or actuator, a microprocessor, memory, a communication module, and a power source. The Lowpower Wireless Personal Area Network (LoWPAN) is a crucial component of the IoT due to its advantageous features such as energy efficiency, widespread accessibility, and the ability to integrate smart objects with the Internet [1].

The IoT has emerged as a transformative force in recent years, connecting billions of devices worldwide and generating massive amounts of data. These connected devices have the potential to drive numerous applications, from smart homes and healthcare to transportation and industry automation.

However, the sheer volume and diversity of IoT data often makes it challenging to extract meaningful insights and enhance user experiences.

One way to address this challenge is by incorporating context awareness into IoT systems. Context-awareness refers to the ability of a system to understand and respond to its environment by considering various contextual factors such as time, location, and user preferences. By incorporating context awareness, IoT devices can adapt to their surroundings and provide personalized experiences to users.

The main contribution in this paper, that we propose a novel Context Awareness IoT framework that combines the Context Awareness concept with the IoT concept, considering performance problems related to limited smart objects resources.

A. Internet of Things

In 1999, Kevin Ashton introduced the concept of the IoT while working at the Auto-ID Center at MIT. The research conducted at this center focused on network radio frequency identification (RFID) and sensor technologies [2], in which People and things could provide information about their current state and their surroundings in a much more efficient manner [3]

IoT comprises wireless systems that are compact in size and interconnected with each other. These systems are equipped with computational capabilities and can transfer data over a network without the need for human interaction. These smart objects are identifiable, can be accessed, and can be programmed locally or remotely via the Internet. They are designed to monitor or control smart spaces. The underutilization of the potential of IoT devices in various domains can be attributed to the limited expressivity and high heterogeneity of the commonly employed scenario programming paradigms [4]. IoT plays a vital role in many domain areas, such as home automation [4], elderly care [5], [6], home safety [7], energy efficiency, and preservation [8], [9].

Recently, research has focused on the connectivity of all physical objects and information environments to the Internet to enable a user to access and control things from anywhere and at any time, which coined the term IoT. IoTs enable

smaller, less complex devices to do complex tasks by enhancing their intelligence and connectivity.

IoT is converting smart spaces from hype into reality. IoT lacks standardization at the application level. IoT devices have limited computational power and memory [10].

The number of devices connected to the Internet has grown significantly in recent years. Perceiving reality through digitizing some parameters of interest can provide a massive amount of data. This data is then shared across the network with other devices, applications, and infrastructures. The IoT paradigm is based on this dynamic and ever-changing world. To date, countless IoT-based applications have been developed considering smart Cities, smart roads, and smart industries [9].

IoT systems face many challenges which make them exploit the intrinsic potential of IoT devices, such as high heterogeneity of the devices and protocols, which results in limited interoperability.

The IoT paradigm has the potential to merge the boundaries between physical objects and computational devices through their interconnectivity via the Internet. This interconnectivity holds the promise of delivering user-centric services that consider both the user's context and profile information [10].

For example, traditional regulation of classroom temperature consists of power on the air conditioner set to the desired temperature and letting the internal thermostat do the rest. However, a smart IoT system could find the best way to cool down the classroom temperature based on the integration of context awareness, ontology, and IoT. For instance, by combining data from internal thermometers, current time, online weather services, and the current number of students, it can decide to just open the window instead of simply turning on the air conditioner, thus saving energy. Also, the system will increase and decrease temperature based on the number of students. Additionally, the same system could automatically close the air conditioner when there is no person in the classroom. Therefore, it is required to manage all forms of data and events that are collected from sensors and devices. For example, simple events (such as the door being opened) and activities (such as the professors and students going out of the classroom) can be translated into context information. It could be accessed by a service that monitors the classroom (e.g., the lecture ended) and a notification sent to the person responsible for this classroom.

B. Context Awareness

Context is defined as information that can be used to characterize the situation of an entity in context-aware computing literature. An entity is a person, location, or thing that is related to the interaction between a user and an application, including the user and the application [11], [12].

On the other hand, context awareness is defined as the property of a system that employs context to provide related information or services to the user, where the relationship is based on the user's task. Therefore, context-aware service discovery can be defined as utilizing context information to discover the most relevant services for the user [11], [13].

Proposing a service-oriented architecture SOA to abstract the complexity in the access to smart devices so that the devices can be viewed as a service (thing as a service). This will enable the smart system development team to focus only on their functional requirements instead of device-specific technical details. Context from the web service perspective:

- From the service requester's perspective, context is defined as the surrounding environment affecting the requester's Web services discovery and access.
- From the perspective of the services, context is defined as the surrounding environment affecting Web services delivery and execution.

To achieve successful IoT systems, there is a need to integrate a context awareness system with IoT.

Context-aware solutions face numerous challenges, including managing the heterogeneous and massive amount of data generated by IoT devices. Second, how to store and handle events, as well as infer higher-level activities from a set of simple events. Finally, use the development framework to implement applications across multiple domains [14].

The growth of the IoT created a fragmented landscape with many devices, technologies, and platforms, creating interoperability issues on many system deployments [15].

Interoperability is one of the most significant barriers to promoting IoT adoption and innovation.

C. Interoperability

Nowadays the ecosystem of the IoT is currently facing a lack in terms of interoperability among the various competing platforms that are presently accessible [16].

Semantic Web (SW) technologies, which consist of an open set of recommendations for associating data with their formal meaning, have been demonstrated to be a suitable means of achieving data interoperability in IoT systems [17]. The reason for using SW technologies is its inference capabilities over semantically annotated data.

Similar to Semantic Web's vision for the Web of Linked Data, the literature on deploying Semantic Web technologies to IoT focuses on semantically annotating data from smart objects. The most prevalent method for representing semantics is Resource Description Framework (RDF), which represents knowledge as triples (subject, predicate, object) (for example, [TempSensor105, Value, 25] and [TempSensor105, Location, Lab2]). A set of triples constitutes a graph consisting of subjects, objects, and predicates. The benefit of RDF and graph data models is that new knowledge can be inferred from an existing graph. Using domain knowledge, a system can comprehend, for instance, that the temperature in Lab 2 is 25 degrees Fahrenheit, which is a transitive property. Web Ontology Language (OWL), one of the primary languages (with RDF schema) used to define ontologies on the web, is frequently used to express domain knowledge to perform the annotation on intelligent [18]. Recently, there has been a lot of research into how SW technologies, in particular OWL ontologies, can be used to improve the IoT field's poor interoperability [4].

Several solutions exist to facilitate interoperability among diverse IoT platforms and application domains. One such solution is the Web of Things (WoT) architecture, recently introduced by the W3C consortium [19].

WoT is a paradigm devised by the World Wide Web Consortium (W3C) on top of the IoT concept. It provides standard mechanisms to interact with any type of device from any automatic system using a descriptive JSON file called Thing Description [20].

Rule-based programming approaches are suitable for IoT automation systems due to their simplicity and intuitive use. One of the most common tools used in programming IoT scenarios using trigger action rules is IFTTT which has many limitations, such as low-level abstraction and low generalization [4], [21].

This paper is organized as follows: an introduction is presented in Section I. Section II illustrates some of the related works. Section III presents the general architecture of the proposed framework. Section IV shows the experiments and results. Section V presents the conclusion and future works.

D. Web of Objects

Web of Objects (WoOs) objectifies and virtualizes real-world objects to support intelligent features and provide Realtime data about the physical world by representing them as Web resources, which can be accessed using the lightweight REST-based APIs principles rather than the heavyweight SOAP-based architecture. Any object with a sensor or actuator, CPU, memory, communication, and power source is considered smart. The web is an ideal universal platform for IoT applications because it uses open standards and can be accessed from any device. In the web environment, sensors or actuators can offer their capabilities via a REST-based API (e.g., URI/lightON and URI/light OFF), which enables objects to interact dynamically. Service provisioning is the process of providing smart object services to the web, similar to traditional web services available on the web. Any web application that communicates with smart objects via communication networks and Web standards is referred to as an IoT application on the web [22].

The smart environment comprises sensors, actuators, interfaces, and appliances networked together to provide localized and remote control of the environment. Sensing and monitoring the environment include temperature, humidity, light, and motion. Environment control such as heater and fan ON/OFF control is provided by the actuator having dedicated hardware interfaces and computing capabilities. Localized control is provided by Bluetooth and remote access through WiFi. The RESTful architecture enables interoperability in Smart space WoOs Architecture.

Semantic ontology helps ubiquitous environments address key issues like knowledge representation, semantic interoperability, and service discovery and provides an efficient platform for building highly responsive and context-aware interactive applications.

According to the following reasons, information systems' ability to communicate with smart objects has become more complicated:

- Many hardware devices rely on proprietary protocols to perform their functions.
- Many devices have embedded software that remains constant over their entire lifespan.
- Semantic annotation for the sensors and the services.
- Service discovery and subscription.
- Simultaneous requests.
- Service authorization.
- Web API generation.

II. LITERATURE REVIEW

There have been several studies focused on how to apply Web paradigms and protocols to service provisioning, such as Service-Oriented Architecture (SOA), RESTful service (REST), Semantic-based provisioning, and the WoT.

Sciullo et al. [15] propose a WoT Store, a centralized repository for managing resources and applications on the WoT. While the proposed system has several potential benefits, there are also some limitations and disadvantages to consider. The WoT Store system may rely heavily on centralized infrastructure, which can lead to scalability, reliability, and security challenges. Additionally, the system may require developers to use a specific set of APIs and communication protocols, which could limit the flexibility and interoperability of IoT devices and applications. Nonetheless, the paper presents a prototype implementation of the WoT Store system and evaluates its performance in terms of resource discovery time and application deployment time, demonstrating the effectiveness of the proposed system for managing IoT resources and applications on the WoT.

Iqbal et al. [22] propose an interoperable IoT platform that can be utilized in a smart home system. The proposed platform employs WoO and cloud architecture. The platform under consideration offers the capability of achieving interoperability among a range of legacy home appliances, diverse communication technologies, and protocols. The platform facilitates remote control of household appliances and enables the storage of home data in the cloud for use by diverse service providers' applications and analytical purposes. The article proposes potential areas for further research, including the incorporation of machine learning algorithms, the deployment of a mobile application, and the creation of a security framework for the smart home system. The proposed architecture is extensible to a variety of smart building use cases, including factories, offices, smart infrastructure, etc.

Ibaseta et al. [23] propose a new methodology for the monitoring and controlling of energy usage in constructions through the utilization of WoT technology. The proposed methodology incorporates diverse building systems and devices, such as lighting, occupancy sensors, and smart plugs, through a cloud-based platform that employs web protocols

and standards. The present study showcases a suggested methodology through a case study of a retrofit project undertaken in an office building. The results indicate that the proposed approach is both cost-effective and energy-efficient while also being interoperable and scalable. Furthermore, the study suggests that this approach can be implemented in real-world settings.

Reda et al. [4] propose a knowledge-based approach for home automation systems using IoT devices. The proposed approach aims to achieve greater expressivity and a higher level of abstraction needed to build knowledge-enabled and reasoning-capable home automation systems so that the potential offered by IoT devices can be fully exploited. The paper demonstrates the feasibility and efficiency of the proposed approach in a simulated house environment, and it contributes to the development of home automation systems by providing a new approach that uses web standards and public ontologies to implement well-defined reasoning without the need for ad hoc control programs or ontologies. The paper also suggests several future works, including investigating the scalability of the proposed approach, evaluating it in a real-life setting, comparing it with other existing approaches, extending it to support more advanced reasoning capabilities, and developing a user-friendly interface for configuring and managing the proposed approach.

Gochhayat et al. [10] propose LISA Lightweight Context-Aware IoT Service Architecture for managing and efficiently managing and delivering push-based services in an IoT environment designed to minimize the overhead of communication and processing while using context information such as device capabilities, user preferences, and environmental conditions to provide personalized and efficient IoT services. LISA filters and forwards the most important and relevant services to the users by understanding their context. The paper evaluates the scalability of LISA through simulations that test its ability to handle many IoT devices and services. The results show that LISA can scale efficiently to support many devices and services while maintaining acceptable levels of performance. However, the paper also acknowledges some limitations of LISA, such as limited support for complex services and limited scalability with centralized deployment.

Krati et al. [24] discuss the challenges of maintaining excellent air quality in indoor environments. It proposes a context aware IoT system that collects data, predicts ventilation conditions, and provides the end user with alerts and recommendations. Through a smartphone application, the system notifies the end-user of contextual information regarding the indoor environment and current ventilation conditions. The system can also provide the end-user with recommendations on improving ventilation and reducing indoor pollutant levels, such as opening windows, installing air purifiers, and modifying the HVAC system. Using the ventilation rate calculated with the aid of interior CO₂ concentration, multilevel logistic regression is used to define indoor ventilation states using ventilation rate. K-NN classification technique to predict ambient ventilation.

Xue et al. [25] examine the advancement of context-aware information fusion technology in the context of smart libraries, employing IoT situational awareness. This paper presents a comprehensive examination of the present status of smart libraries and context-aware technology while also conducting an analysis of the implementation of context-aware technology within smart libraries. This paper presents a conceptual framework for implementing a smart library, which leverages context awareness to enhance its services. It further proposes integrating context-aware services within smart libraries to optimize user experiences and improve overall functionality. The aforementioned findings propose potential avenues for future research in this field. These include the advancement of more precise algorithms for context-aware information fusion, the investigation of alternative IoT technologies, and the resolution of ethical and privacy issues. In general, the paper emphasizes the potential of the IoT situational awareness technology in improving the efficacy and efficiency of intelligent libraries.

Kim et al. [26] present a middleware architecture for a context-aware system in a smart home environment. To infer high-level contexts from available low-level contexts, the proposed architecture incorporates a profile-applied improved rule-based reasoning algorithm. The context is modeled using OWL and ontology. The experimental result demonstrates that the middleware provides more accurate and quicker reasoning results than the conventional rule-based method. In addition, the context-aware service is also selected using a rule-based algorithm, allowing the service to be readily expanded by adding new service rules to the service rule base.

In this paper, a Service-Oriented Context-Aware Middleware (SOCAM) architecture is proposed for developing and prototyping context-aware services in pervasive computing environments. To develop context-aware services, architecture provides efficient support for acquiring, discovering, interpreting, and accessing diverse contexts. In addition, the paper proposes a formal context model based on Web Ontology Language and ontology to address issues such as semantic representation, context reasoning, context classification, and dependency. The context model and middleware architecture are described, as well as the prototype's performance in a smart home environment [27].

III. PROPOSED FRAMEWORK

IoT service provisioning techniques must take into consideration how to provide service consumers with complete data, including sensor data as well as its context and don't overwhelm the consumer with repeated information.

In addition, the rapid growth of IoT services available to users will result in a significant increase in information overload and network resource consumption. The exponential growth in the number of services available presents a challenge in selecting and providing the appropriate service from the vast array of required services. Therefore, to locate the most pertinent service, it is essential to construct an accurate query that considers the consumer's context.

In this section, we present our proposed service provisioning framework called CSSP, Fig. 1 presents the proposed framework, and we will discuss how to overcome the problems related to integrating IoT and Context-aware systems in a scalable manner which will be described in detail in the following sections.

A. Data Acquisition and Management Service

This layer is the core layer of the framework which is responsible for all operations related to collecting sensor data and storing it locally. This layer has many data collection mechanisms to overcome the limitation of IoT sensors. The purpose of this layer is to hide the details of data collection and protocols used from users; user may be a web developer or

mobile developer who interacts with sensors as a virtual object. Our methodology for collecting sensors data is based on three levels:

- Collect sensor data using the suitable protocol and according to the user role (Fig. 2 shows accessing sensor value through CoAP protocol in the Cooja simulator from a web browser).
- Annotate sensor data with context using ontology. For example, Fig. 3 shows how to model the value of temperature using context aware. There are many methods to represent RDF, such as Triples, Shorthand notation, and XML notation, showing the use of Shorthand notation code to model sensor values.

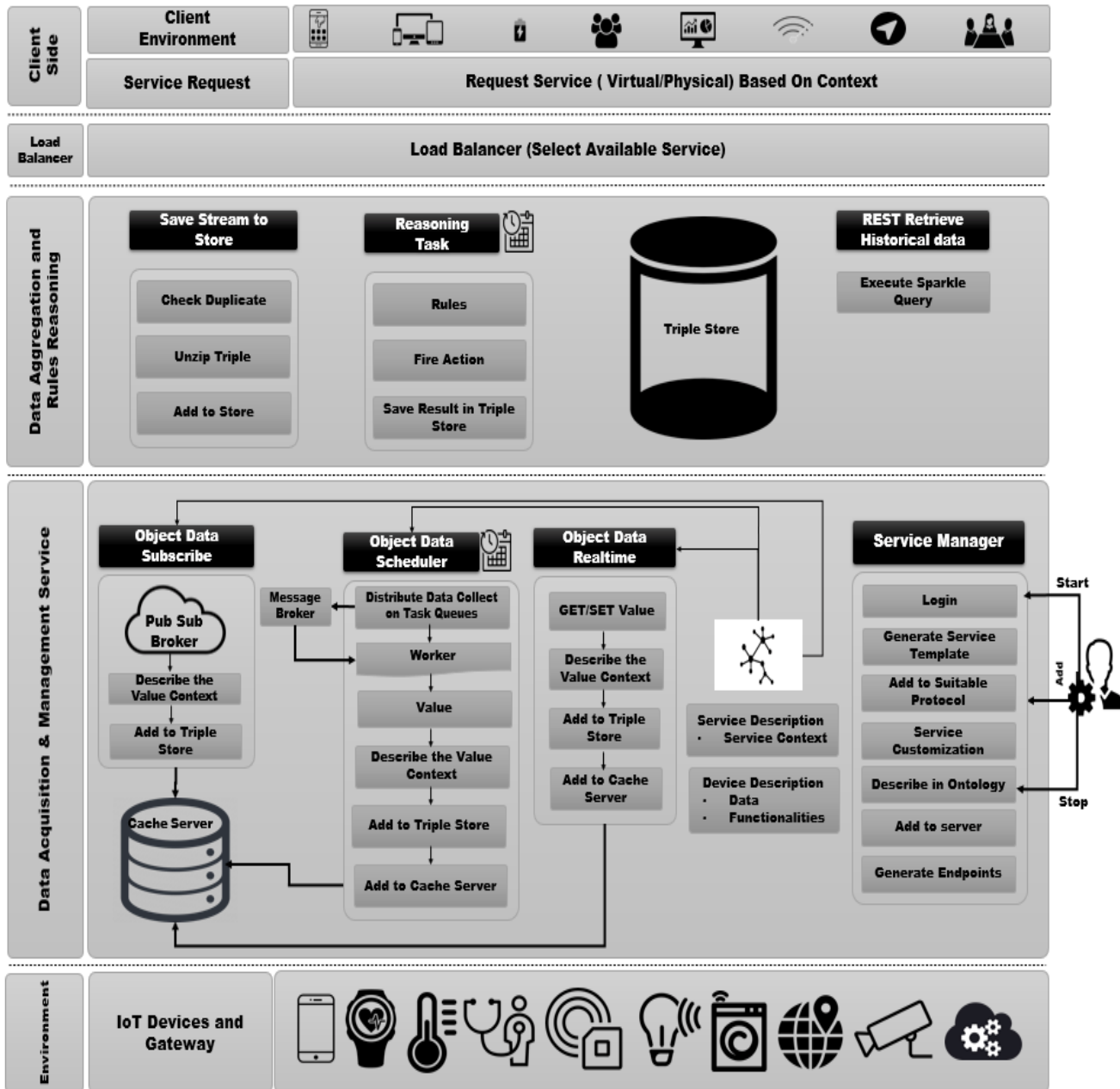


Fig. 1. The proposed framework (CSSP).

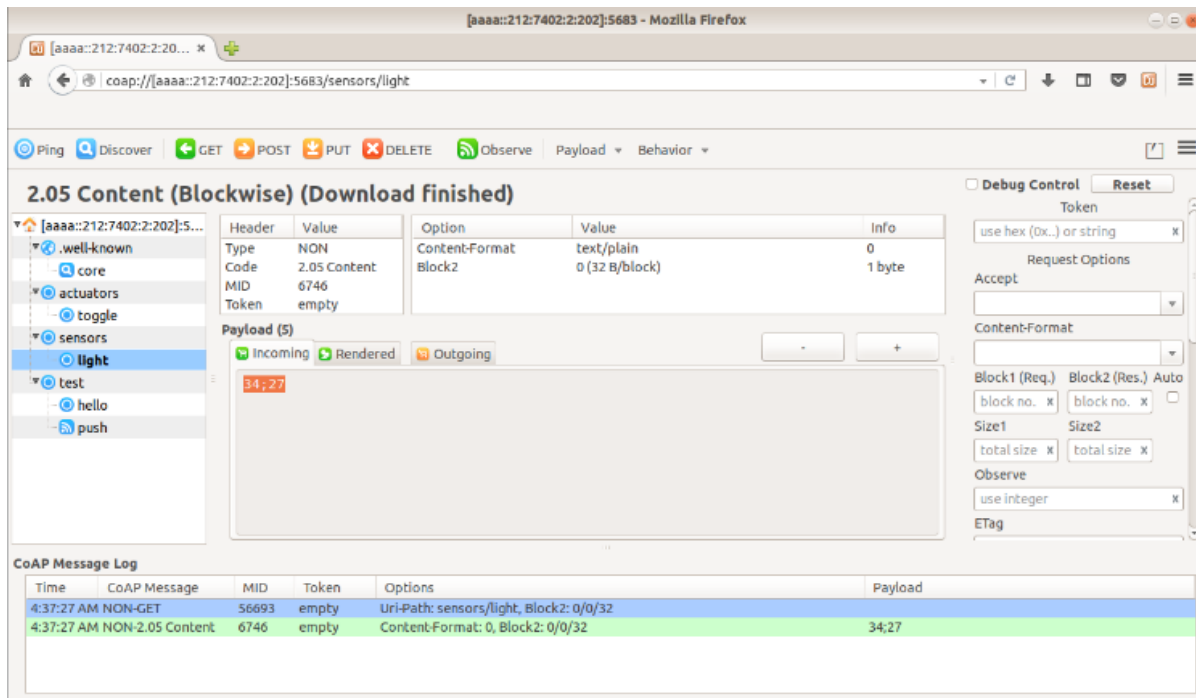


Fig. 2. Accessing CoAP sensor value.

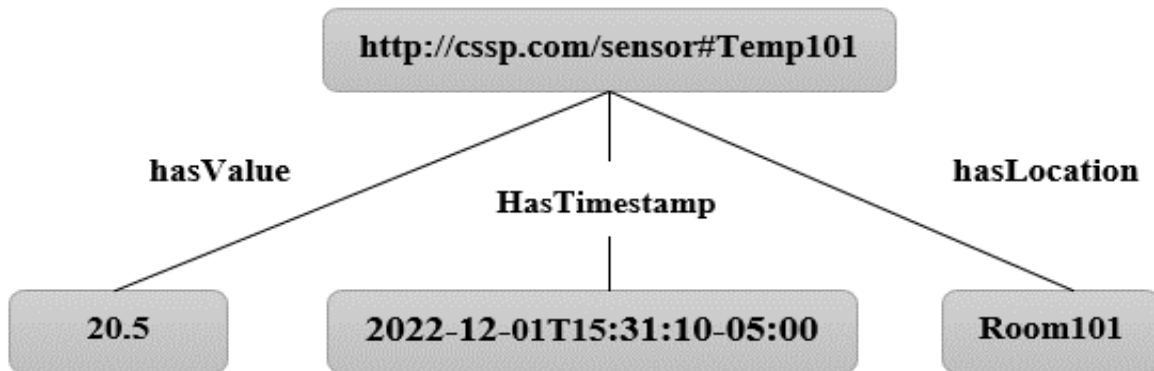


Fig. 3. Temperature sensor value representation using RDF Graph.

The example in Listing. 1 show how to use shorthand notation to model sensor data.

```

@prefix so: <http://cssp.com/sensor#> .
so:Temp101 hasValue 20
so:Temp101 hasTimestamp 2022-12-10T15:31:00Z^^xsd:dateTime
so:Temp101 hasLocation Room101
  
```

Listing. 1. Sensors reading representation using RDF

- Save the generated data in a triple store. This component starts working through the system admin when the admin wants to add a sensor. The admin fills in some data about the sensor, and the component generates an abstract service template according to the specified protocol. The admin customizes the service, such as customizing the result format or any metadata about the result, and how to read the data from pin id, etc.

The service itself is described using an ontology. Every detail about the service is described, such as URL, Endpoints, data format, category, location, etc. Admin can start, stop, and remove any service in the framework at any time. After the admin creates, customizes, and describes the service, the service will be added to the server to allow access to sensor data based on identified mechanism.

In our view of data Acquisition of sensors, it depends on the used protocol, the role of the user, and the status of data (Realtime data or cached instant data).

We can view the data generated from IoT sensors as the following:

- Object Data Realtime
- Object Data Scheduler
- Object Data Subscribe

All these components depend on ontology to describe IoT devices.

1) *Object data realtime*: This is the simplest and most direct way to access sensor data which is used by Realtime role users. Based on the type of IoT device, it offers GET operation in case of getting data from the device, else if the device is an actuator, it will have GET and POST operations, Also Mapping of CoAP [28] URI requests to standard web requests.

Data generated from sensors or written to actuators are not meaningful without a description, so we use ontology to describe the context of sensor values such as location, time, and available people.

After the description of sensor data, we add this data to triple store locally and the most recent value to a cache server which can be used by other users who are not concerned with instant data from sensors.

This role is very important in cases of emergencies where milliseconds can be important for patients or the elderly, so we must decrease the request count on sensors for any user.

2) *Object data scheduler*: This is one of the most important components of the framework and is commonly used by many types of users. This mechanism works with the normal users of the system.

There are many task queues to distribute task requests and categorize task queues for every type of sensor. First, it checks the cache server to see if available recent data is present. If ok, it will return and delete the request. Else, the request will be added to the suitable task queue. After the worker gets the value, the component describes the value, and it is to triple-store and adds the value to the cache server Scheduler threshold settings are configured using system admin.

3) *Object data subscribe*: This mechanic is frequently used for monitoring purposes, and it uses different protocols, such as Message Queuing Telemetry Transport MQTT [28] to allow the user to subscribe to a collection of sensor value changes. The layer also adds context data to the value returned and adds it to the triple store and cache server.

B. Data Aggregation and Rules Reasoning

In the Data Acquisition and Management Service layer, the data is distributed on edges and have duplicates in the data aggregation layer. The framework will do the following:

- Save Stream to Store
- Reasoning Task
- Retrieve Historical Data

1) *Save stream to store*: In this phase, First, the system checks sensors data and stores only incoming new unique data and ignores repeated values. The service receives only new values to reduce the size of the data. Secondly, extracted common data from ontology was done on the local edges. In this phase, we complete the data of the sensor. We extract

metadata or minimize to reduce used bandwidth used to transfer sensor data from edges to a data store. Finally, the IoT value will be stored in the triple store database.

2) *Reasoning task*: In this phase, we define rules which can be used to infer new knowledge from sensor data. Any rules engine base can be used, such as SWRL rules. When any activity is detected based on rules, sensors, and context data, action is taken and also described and saved in the RDF triple store. The following example in Listing. 2 demonstrates an example of the SWRL rule to regulate air conditioners in a classroom based on current readings from temperature sensors, the number of persons in the same room, and the state of the window.

In this example, `hasNumberOfStudents`, `hasTemperature`, and `hasWindowState` are individual properties that represent the number of students in the classroom, the temperature in the room, and the state of the window (open or closed), respectively. `hasAirConditionerState` and `hasTemperatureSetting` are individual properties that represent the state of the air conditioning system (on or off) and the temperature setting, respectively. The SWRL rules in this example take into account the number of students in the room, the temperature in the room, and the state of the window to determine the appropriate state and temperature setting for the air conditioning system. For example, if there are less than 10 students in the room and the temperature is above 25°C, the air conditioning system will be turned on and set to cool the room down to 24°C. On the other hand, if the window is open, the air conditioning system will be turned off regardless of the number of students or temperature.

3) *Retrieve historical data*: Cause we have a triple store, we can run the system in two methods:

- REST API: this is a simple method and can be used by any developer to call API for getting device data by its id and during a specific period.
- SPARQL Query: this is the standard language used to query from triple store TDB and which will be important in the case of a query for historical data in the TDB.

C. Client Side

The client-side layer is the layer responsible for collecting context data and sending this context with the service request, for example, when the user in a smart home environment wants to turn on the air conditioner, and there are many types of contexts information that can be sent with requests such as the room of the user (location), current temperature from internet service, number of users in the rooms, and preferred temperature from user profile.

D. Load Balancer

The load balancer layer is an optional layer in our framework according to the application size and number of users. This layer is responsible for receiving a request, then finding an available server and routes the request to this server. Load balancers, including physical appliances, software instances, or a combination of the two.

```
hasNumberOfStudents(?classroom, ?numStudents) ∧ hasTemperature(?classroom, ?temp) ∧ hasWindowState(?window, ?state)
  → hasAirConditionerState(?aircon, ?airconState)
∧ lessThan(?numStudents, 10) ∧ greaterThan(?temp, 25) ∧ equal(?state, closed)
  → hasAirConditionerState(?aircon, on)
∧ hasTemperatureSetting(?aircon, 24) ∧ greaterThanOrEqualTo(?numStudents, 10)
∧ lessThanOrEqualTo(?numStudents, 20) ∧ greaterThan(?temp, 25) ∧ equal(?state, closed)
  → hasAirConditionerState(?aircon, on)
∧ hasTemperatureSetting(?aircon, 23) ∧ greaterThan(?numStudents, 20) ∧ greaterThan(?temp, 25) ∧ equal(?state, closed)
  → hasAirConditionerState(?aircon, on) ∧ hasTemperatureSetting(?aircon, 22)
∧ equal(?state, open)
  → hasAirConditionerState(?aircon, off)
```

Listing. 2. the SWRL rule example to regulate air conditioners in classroom

E. Environment

This layer is responsible for the IoT environment we want to monitor and control locally or through the Internet, which has sensors and actuators. This layer needs to use a gateway such as Arduino and Raspberry Pi to control various digital or analog devices.

IV. EXPERIMENTS

To confirm the proposed framework's effectiveness, an experiment has been applied to evaluate the response time in the case of traditional requests and in the case of our framework. We build a simple IoT network in a Cooja simulator [29] and test simultaneous requests for the sensors, then measure and compare the performance of requests for the traditional request and our framework.

1) *Dataset*: The environment used to generate these datasets was Contiki 3.0 OS (Ubuntu 18.04 Based) in a virtualization environment (VMware application). The device used for generating datasets is a laptop with a Core I5 processor and 8 GB Ram (2 GB for the virtual machine). The Cooja simulator was used to build a simulation of an IoT network.

We built a Python application to simulate the request of the huge simulation for a sensor in case of a CoAP request and proposed HTTP request (500 requests simultaneously) and then saved the details of requests and responses in a CSV file. In the case of a CoAP request, the log file will contain the following attributes: request id, start time, end time, response time, status, and value. In the case of the proposed HTTP, we added an attribute to indicate the source of data, which will be physical sensor data or from a cache server. Two data sets for CoAP and HTTP are generated.

2) *Results and discussion*: Based on the simulation results, it is observed that an increased number of CoAP requests to the sensor resulted in a significant exponential increase in response time. This effect can be attributed to the absence of scheduling or caching mechanisms, as the direct access of sensor values contributed to longer response times. A sample of requests and their corresponding response times are illustrated in Table I (Sample of CoAP Requests Dataset) and Table II (Sample of Proposed HTTP Requests Dataset).

TABLE I. SAMPLE OF COAP REQUESTS DATASET

Request Number	Response Time
145	1.9
404	2.6
147	3.8
474	4.6
48	6.9
281	7.9
70	8.9
442	9.9
419	10.5
41	14.9
112	16.9
78	17.9
126	18.8
289	19.9
204	20.9
346	21.9
475	22.2
98	31.9
250	32.8

TABLE II. SAMPLE OF PROPOSED HTTP REQUESTS DATASET

* C: Cached		* R: Realtime	
Request ID	Response Time	Value	Type
73	0.11	240	C
132	0.15	240	C
347	0.14	76	C
1	0.72	240	R
38	2.16	240	C
364	3.37	122	C
252	4.74	122	C
95	7.13	122	C
10	15.53	54	L

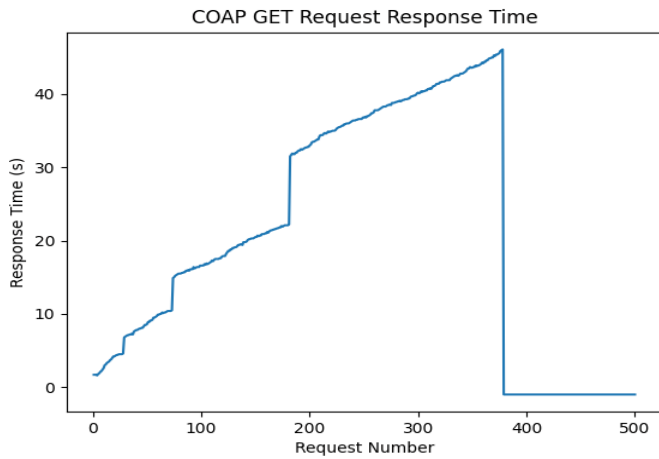


Fig. 4. CoAP response time.

In the traditional approach, the results from our analysis showed that approximately 32% of the requests failed due to this lack of optimization and huge simultaneous requests to the same sensor. The processing time for all requests is 47 seconds. Fig. 4 shows a graph representing the sensors direct access performance through the CoAP protocol.

The implementation of scheduling and caching mechanisms as well as a time threshold established in the system configuration affect the response time in our proposed prototype. As per this threshold, the initial request may experience a relatively longer latency, but subsequent requests are expected to have faster and more consistent response times.

In the proposed approach, the results from our analysis show that the simulation conducted with this prototype revealed a failure rate of 0%. Additionally, the data analysis in the table indicates that a significant portion of sensor data is obtained from the cache server. Consequently, most responses were delivered within an acceptable timeframe, ensuring efficient and reliable system performance. The processing time for all requests is 16 seconds. Fig. 5 shows a graph representing the sensors access performance through the proposed framework based on the HTTP protocol.

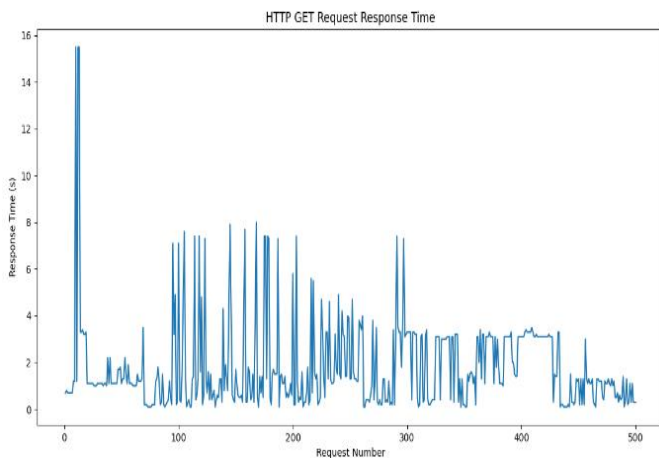


Fig. 5. Poposed HTTP response time.

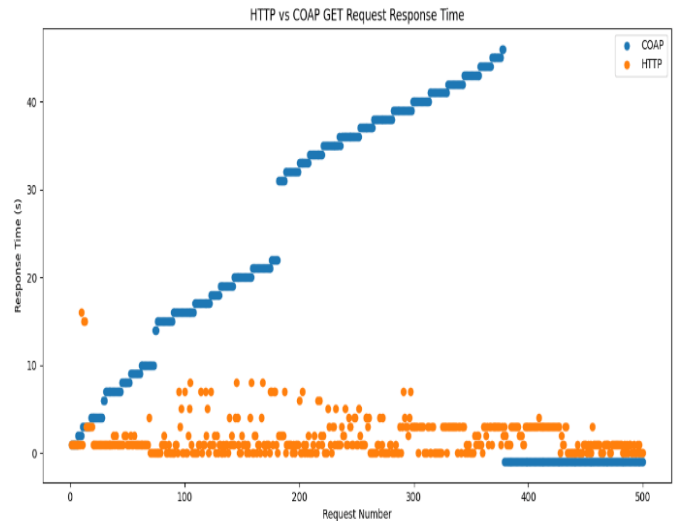


Fig. 6. Performance comparison of CoAP and proposed HTTP.

Fig. 6 shows a graph representing a comparison of performance according to response time to the CoAP protocol (traditional direct access) and HTTP protocol (Proposed framework).

V. CONCLUSION AND FUTURE WORKS

This paper proposes a context-aware semantic service provisioning framework. The framework focuses on producing solutions for many problems by providing smart devices as smart objects in standard web environments and facilitating the development of any category of IoT application.

The framework was organized to solve many network problems such as bandwidth, scalability, limited resources of smart devices, semantic annotation, and reasoning.

The proposed framework is validated and evaluated by testing accessing the service in the traditional CoAP protocol and according to the proposed framework and comparing the results. The evaluation shows that the proposed framework increases performance and decreases failed requests.

In the future, we will try to test this framework in a real environment, and we will try to increase the performance.

REFERENCES

- [1] S. N. Han and N. Crespi, "Semantic service provisioning for smart objects: Integrating IoT applications into the web," *Future Generation Computer Systems*, vol. 76, pp. 180–197, Nov. 2017, doi: 10.1016/J.FUTURE.2016.12.037.
- [2] K. Avila, P. Sanmartin, D. Jabba, and M. Jimeno, "Applications Based on Service-Oriented Architecture (SOA) in the Field of Home Healthcare," *Sensors (Basel)*, vol. 17, no. 8, Aug. 2017, doi: 10.3390/S17081703.
- [3] Kevin Ashton, "That 'internet of things' thing," *RFID journal*, vol. 22, no. 7, 2009.
- [4] R. Reda et al., "Supporting Smart Home Scenarios Using OWL and SWRL Rules," *Sensors (Basel)*, vol. 22, no. 11, Jun. 2022, doi: 10.3390/S22114131.
- [5] S. Y. Y. Tun, S. Madanian, and F. Mirza, "Internet of things (IoT) applications for elderly care: a reflective review," *Aging Clin Exp Res*, vol. 33, no. 4, pp. 855–867, Apr. 2021, doi: 10.1007/S40520-020-01545-9/TABLES/4.

- [6] A. S. Salama and A. M. Eassa, "IOT AND CLOUD BASED BLOCKCHAIN MODEL FOR COVID-19 INFECTION SPREAD CONTROL," *J Theor Appl Inf Technol*, vol. 15, no. 1, 2022, Accessed: Aug. 22, 2023. [Online]. Available: www.jatit.org
- [7] Taryudi, D. B. Adriano, and W. A. Ciptoning Budi, "Iot-based Integrated Home Security and Monitoring System," *J Phys Conf Ser*, vol. 1140, no. 1, p. 012006, Dec. 2018, doi: 10.1088/1742-6596/1140/1/012006.
- [8] V. Marinakis and H. Doukas, "An Advanced IoT-based System for Intelligent Energy Management in Buildings," *Sensors (Basel)*, vol. 18, no. 2, Feb. 2018, doi: 10.3390/S18020610.
- [9] M. Lombardi, F. Pascale, and D. Santaniello, "Internet of Things: A General Overview between Architectures, Protocols and Applications," *Information 2021*, Vol. 12, Page 87, vol. 12, no. 2, p. 87, Feb. 2021, doi: 10.3390/INFO12020087.
- [10] S. P. Gochhayat et al., "LISA: Lightweight context-aware IoT service architecture," *J Clean Prod*, vol. 212, pp. 1345–1356, Mar. 2019, doi: 10.1016/J.JCLEPRO.2018.12.096.
- [11] V. Ponce and B. Abdulrazak, "Context-Aware End-User Development Review," *Applied Sciences 2022*, Vol. 12, Page 479, vol. 12, no. 1, p. 479, Jan. 2022, doi: 10.3390/APP12010479.
- [12] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggle, "Towards a better understanding of context and context-awareness," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 1707, pp. 304–307, 1999, doi: 10.1007/3-540-48157-5_29/COVER.
- [13] S. A. Z. Hassan and A. M. Eassa, "A Proposed Architecture for Smart Home Systems Based on IoT, Context-awareness and Cloud Computing," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, pp. 89–96, Autumn 2022, doi: 10.14569/IJACSA.2022.0130612.
- [14] M. Elkady, A. Elkorany, and A. Allam, "ACAIOT: A framework for adaptable context-aware IoT applications," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 4, pp. 271–282, 2020, doi: 10.22266/IJIES2020.0831.24.
- [15] L. Sciuillo, L. Gigli, A. Trotta, and M. Di Felice, "WoT Store: Managing resources and applications on the web of things," *Internet of Things*, vol. 9, p. 100164, Mar. 2020, doi: 10.1016/J.IOT.2020.100164.
- [16] J. Lanza, L. Sánchez, D. Gómez, J. R. Santana, and P. Sotres, "A Semantic-Enabled Platform for Realizing an Interoperable Web of Things," *Sensors (Basel)*, vol. 19, no. 4, Feb. 2019, doi: 10.3390/S19040869.
- [17] D. Andročec, M. Novak, and D. Oreški, "Using Semantic Web for Internet of Things Interoperability," *Int J Semant Web Inf Syst*, vol. 14, no. 4, pp. 147–171, Oct. 2018, doi: 10.4018/IJSWIS.2018100108.
- [18] F. Z. Amara, M. Hemam, M. Djezzar, and M. Maimour, "Semantic Web Technologies for Internet of Things Semantic Interoperability," *Lecture Notes in Networks and Systems*, vol. 357 LNNS, pp. 133–143, 2022, doi: 10.1007/978-3-030-91738-8_13/COVER.
- [19] "Solution for IoT Interoperability - W3C Web of Things (WoT)." <https://www.w3.org/2020/04/pressrelease-wot-rec.html.en> (accessed Apr. 01, 2023).
- [20] S. Murawat et al., "WoT Communication Protocol Security and Privacy Issues," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 155–161, 2020, doi: 10.14569/IJACSA.2020.0110319.
- [21] "IFTTT - Connect Your Apps." <https://ifttt.com/> (accessed Dec. 18, 2022).
- [22] A. Iqbal et al., "Interoperable Internet-of-Things platform for smart home system using Web-of-Objects and cloud," *Sustain Cities Soc*, vol. 38, pp. 636–646, Apr. 2018, doi: 10.1016/J.SCS.2018.01.044.
- [23] D. Ibaseta et al., "Monitoring and control of energy consumption in buildings using WoT: A novel approach for smart retrofit," *Sustain Cities Soc*, vol. 65, p. 102637, Feb. 2021, doi: 10.1016/J.SCS.2020.102637.
- [24] K. Rastogi, D. Lohani, and D. Acharya, "Context-Aware Monitoring and Control of Ventilation Rate in Indoor Environments Using Internet of Things," *IEEE Internet Things J*, vol. 8, no. 11, pp. 9257–9267, Jun. 2021, doi: 10.1109/JIOT.2021.3057919.
- [25] X. Chen and Q. Hao, "Research on Internet of Things Context-Aware Information Fusion Technology for Smart Libraries," *Sci Program*, vol. 2022, 2022, doi: 10.1155/2022/5282932.
- [26] H. W. Kim, M. R. Hoque, H. Seo, and S. H. Yang, "Development of middleware architecture to realize context-aware service in smart home environment," *Computer Science and Information Systems*, vol. 13, no. 2, pp. 427–452, Jun. 2016, doi: 10.2298/CSIS150701010H.
- [27] T. Gu, H. K. Pung, and D. Q. Zhang, "A service - oriented middleware for building context - aware services," *Journal of Network and Computer Applications*, vol. 28, no. 1, pp. 1 - 18, Jan. 2005, doi: 10.1016/J.JNCA.2004.06.002.
- [28] S. Bansal and D. Kumar, "IoT Ecosystem: A Survey on Devices, Gateways, Operating Systems, Middleware and Communication," *Int J Wirel Inf Netw*, vol. 27, no. 3, pp. 340–364, Sep. 2020, doi: 10.1007/S10776-020-00483-7/FIGURES/7.
- [29] G. Oikonomou, S. Duquenooy, A. Elsts, J. Eriksson, Y. Tanaka, and N. Tsiftes, "The Contiki-NG open source operating system for next generation IoT devices," *SoftwareX*, vol. 18, p. 101089, Jun. 2022, doi: 10.1016/J.SOFTX.2022.101089.

Impact of the Use of the Video Game SimCity on the Development of Critical Thinking in Students: A Quantitative Experimental Approach

Jorge Luis Torres-Loayza, Grunilda Telma Reymer-Morales and Benjamín Maraza-Quispe
Facultad de Ciencias de la Educación de la, Universidad Nacional de San Agustín, de Arequipa-Perú

Abstract—The objective of the research is to determine to what extent the use of the SimCity video game allows the development of critical thinking in the teaching-learning processes of students. The methodology applied consisted of a research with a quantitative approach of experimental type, working with a sample of 25 students selected through a simple random sampling of a population of 100 students, 10 sessions were developed using the SimCity video game, a pretest and posttest of skills and abilities required to develop critical thinking of Watson Glaser were applied, whose dimensions measured were: Inferences, assumptions, deductive reasoning, logical interpretation and evaluation of arguments. The results show that with adequate stimulation through the use of the SimCity video game, critical thinking can have a moderate but effective development in the students, from the comparison of the data obtained in the pretest and posttest a significant progress in terms of scores is observed; likewise, the effectiveness of the use of the SimCity video game is reflected to a greater extent in inferences and evaluations of arguments, since during the posttest evaluations greater progress was observed in comparison to other skills; while the interpretation of information obtained less progress in comparison to the other skills, the use of skills such as deductive reasoning, inferences and evaluation of arguments were moderately developed. In conclusion, the use of the SimCity video game allows the development of skills and abilities to develop critical thinking according to various factors, such as the way in which it is incorporated into the curriculum, the orientation and guidance of teachers, and the way in which reflection and analysis is carried out after the game experience.

Keywords—*SimCity; video games; critical thinking; critical learning*

I. INTRODUCTION

In many educational institutions, the teaching approach focuses on memorization and repetition of information rather than critical analysis of it. This can hinder the development of critical thinking in students as they are not given the opportunity to question and critically evaluate information. In this context, we propose the following research question: To what extent does the use of the video game SimCity foster the development of critical thinking in students?

SimCity is a game designed for educational and critical decision-making purposes. The study [1] mentions that these games, besides being created by a dedicated team of game developers, also involve psychologists, doctors, and other professionals who assist in creating a more serious and realistic environment. Thus, SimCity is a serious simulation game

where decisions depend on the player. According to [2], a video game can be a simulation game, which, unlike a simulator, tends to have open gameplay based on strategy and role-playing, as opposed to a rigid action and adventure gameplay of a simulator. SimCity is a video game software created by Electronic Arts and released in 2003. The game is designed for single player, but there is an option to create multiple cities that can be played by different people at different times within the same software. It is a system simulation game where players assume the role of a mayor and deal with issues such as crime, budget deficits, and traffic to create and develop cities [3]. Additionally, [2] mentions that as a simulator, one must think as if everything were real and consider all the factors present since the player holds the power. The simulation game SimCity offers the opportunity to orchestrate the construction and development of a city. The tremendous success of SimCity demonstrates the surprisingly convincing power of a particular type of human-computer interaction. According to [3], in combination, it can be understood that it is a popular game that allows for interaction between people and machines. The success of the game in education is due to the artificial environment with rules where there are no limits of size or time [4]. Since the game presents a large amount of data that players must be aware of, such as air pollution levels. SimCity 4 allows its players to shape and build a settlement where the player acts as the mayor [5]. SimCity allows players to determine the evolution of a city, which depends on the player's decisions. Thus, SimCity has an interface with a small control panel that provides information about the city's development and any problems it may have, whether they are economic, social, and etc.

The SimCity game guide mentions that decision-making is the main aspect, as players, in the role of mayor, must define zones for houses, factories, and others while also meeting basic needs in order to maintain a stable city. At the beginning, it is recommended to follow a small tutorial to understand the game's themes [6]. The city is not just a toy that SimCity claims to be, and SimCity plays with urban planning in a way that misleads students in their understanding of how a city works and how it could be planned [3]. SimCity is considered a serious game [7]. They mention: Games have the unique ability to provide participants, especially learners, with the opportunity to acquire skills through activities embedded within the game itself. This is due to the playful and interactive nature of games, which allows players to experience situations and challenges actively and participatively. Unlike

gamification, where game elements are incorporated into contexts that are not necessarily playful, games themselves are inherently playful by design. In other words, when it is mentioned that games allow for the acquisition of skills through game-based activities, it refers to how players learn and enhance their abilities as they interact with the game. These activities can involve solving puzzles, making strategic decisions, overcoming obstacles, and achieving goals within the game. Through these actions, players are exposed to challenges that require the application of specific skills, thus contributing to the acquisition of skills.

According to [8], thinking is defined as the ability to process information and build knowledge through the combination of mental representations, operations, and attitudes. It classifies thinking into three types:

- Automatic Thinking: Sometimes we act without much thought, we think automatically; that is, we respond immediately to various stimuli in the environment with previously learned responses.
- Systematic Thinking: Other times, we stop to think, we think systematically; we use all the intellectual resources at our disposal (concepts, skills, and attitudes) to create new responses to situations.
- Critical Thinking: Finally, in very extraordinary occasions, we reflect on our own thinking process; we carry out what philosophers call self-awareness and psychologists call metacognition. This involves examining and critically evaluating our own thoughts, beliefs, and assumptions, as well as the arguments and evidence supporting our conclusions. Critical thinking aims to objectively analyze, question, and rigorously evaluate information before reaching conclusions or making decisions.

Furthermore, it is mentioned that the stage of formal operations, also known as the fourth stage, is the highest point at which a person can make decisions through broader logic [9]. Thus, it can be understood that learning theory is related to critical thinking since in its fourth stage, conclusions are sought through more exhaustive reasoning.

Critical Thinking is configured as a philosophical type of thinking that is concerned with cultivating and improving each individual's reasoning abilities [10]. In other words, it is the capacity that a person possesses to have a critical judgment in a given situation. Critical thinking is a mode of thinking about any topic, content, or problem in which the thinker improves the quality of their thinking by grasping the inherent structures of the act of thinking and subjecting them to intellectual standards [11]. In other words, critical thinking is a cognitive skill that involves analyzing information objectively and systematically, carefully evaluating it, and making informed decisions based on available evidence. Critical thinking involves the ability to analyze, synthesize, and evaluate information from different sources, including books, articles, websites, media, and personal experience.

The Multidimensionality of Critical Thinking has been categorized by [12] as a mode of thinking that corresponds to

the current conditions of the development of productive and technological forces, their complexity, and the changes underlying the multidimensionality of systems. Thus, it is understood that critical thinking seeks to relate all possible aspects of a topic to understand their influence on it. Critical thinking constitutes a complex system of dimensions that allows for the analysis of one's own thinking and that of others at a more specific level, which, when subjected to the analysis of the different dimensions, as stated by [13].

- Logic: Involves analysis through logical reasoning.
- Substantive: Evaluates truth and falsehood.
- Contextual: Recognizes the social and historical context.
- Dialogical: Examines thoughts in relation to the perspectives of others.
- Pragmatic: Recognizes the purpose or intention behind a thought.

It has been researched that SimCity can be applied in the development of administrative skills in undergraduate and graduate students [14]. This indicates that SimCity has application in the educational field. Additionally, Piaget's learning theory reflects the fourth stage when playing a strategy video game, as the player will use logic to make decisions [15]. Thus, SimCity, being a strategy game where decisions must be made through structured logic, allows for the development of critical thinking. Unlike many games, SimCity encourages deep thinking due to its characteristics. Of course, no matter how much computer game designers grant to players, any simulation will be based on a set of basic assumptions. SimCity has been criticized from both the left and the right for its economic model [3]. Therefore, individuals who play SimCity must apply their knowledge based on the dimensions of critical thinking. In strategy games, it is mentioned that logical thinking and problem-solving are primarily developed [16]. Thus, it can be understood that SimCity as a game allows for the development of logic as another dimension of critical thinking, but as a simulator, other aspects must also be considered. Furthermore, [17] mentions that the game allows for decision-making based on the environmental conditions displayed during the construction and development of the city. Therefore, to make decisions, the contextual and dialogical dimensions are important because one has to think based on what is proposed during the game and then make decisions.

On the other hand, the use of algorithms is common in video games. The research [18] mentions how video games nowadays use predictable algorithms, but with the use of artificial intelligence, this varies, although they become more comprehensible and logical. Thus, the pragmatic dimension will also be developed as, during the game, one will have to anticipate and think about the future critically and question the decisions made.

Therefore, critical thinking is essential in decision-making to make them with determination. The author in [19] mentions that video games improve reasoning and critical thinking abilities, leading to making accurate decisions.

According to the research conducted by [20], there is an effect of the game SimCity on the development of spatial intelligence among students at Angkasa Secondary School in Bandung. Secondly, the application of the SimCity game in geography learning helps students make decisions to overcome spatial problems. This is evidenced by the increase in the average score in the post-test of the experimental class after using SimCity as an instructional medium. Thirdly, the application of the SimCity game in geography learning requires structured learning scenario planning and proper time allocation. Consequently, the learning steps must be carefully followed without skipping any steps. Fourthly, SimCity game can be an alternative learning medium, although it has some disadvantages compared to conventional methods.

Likewise, the research conducted by [21] demonstrated that even a technically complex and fast-paced medium like a computer video game, SimCity, can be used in an instructional scenario for an extended period with limited effort, although the challenge of the game's aging must be actively addressed. Furthermore, in the described instructional scenario, SimCity is perceived as a motivating and accepted learning diversification: the learning activity can be considered playful. This finding, moreover, has not been widely presented in the literature.

II. RESEARCH METHODOLOGY

The applied methodology follows a quantitative experimental approach.

A. Objective

To determine to what extent the use of the video game SimCity in teaching-learning processes allows the development of skills and abilities to promote critical thinking in regular basic education students.

B. Hypothesis

The use of the video game SimCity in teaching-learning processes enables regular basic education students to develop skills and abilities to promote critical thinking.

C. Variables

- Independent: Use of the video game SimCity
- Dependent: Development of critical thinking
- Controlled: The number of students and the playtime of SimCity

D. Population and Sample

A simple random sampling has been applied to select 25 students for the control group and 25 students for the experimental group. The total population consisted of 100 third-grade students from regular secondary education. Informed consent was obtained from the parents or guardians of all students.

E. Methodological Design

The research is experimental in nature. According to [22], unlike theoretical research, it requires direct interaction with the members and components of the group being studied, in this case, individuals. To determine the effectiveness of the

video game SimCity in the development of critical thinking, the Watson Glaser Critical Thinking Appraisal test [23] will be administered. This test is internationally standardized.

F. Aspects Measured by the Watson Glaser Test

- Capacity for critical thinking.
- General understanding of the importance of providing evidence and support when formulating conclusions.
- Ability to differentiate between inferences, assumptions, and generalizations through the application of logic.
- The ability to combine these skills in decision-making.

Table I shows that both Group A and Group B participated in the study and were administered the Watson Glaser Critical Thinking Appraisal test in both the pretest and posttest. Additionally, a total of 10 sessions of 30 minutes each were conducted as part of the study. The total duration of the study was 12 hours.

TABLE I. PROCEDURE FOLLOWED IN THE RESEARCH

Group	Pretest	Sessions	Posttest	Duration
Group A and B	The Watson Glaser Critical Thinking Appraisal test will be administered to both samples	10 sessions of 30 minutes each will be conducted	The Watson Glaser Critical Thinking Appraisal test will be administered to both samples	12 hours

G. Application of the Watson Glaser Test to Both Samples

The Table II displays the distribution of questions and scores across different dimensions of critical thinking.

TABLE II. STRUCTURE OF THE WATSON GLASER TEST

Dimensions	Number of questions	Score
Inferences	15	15
Assumptions	16	16
Deductive Reasoning	9	9
Logical Interpretation	12	12
Evaluation of Arguments	15	15
Total	67	67

H. Sessions for the use of the Video Game SimCity

1) Introduction to the game SimCity

- In a session of 25 to 30 minutes, the students will be explained about the game SimCity and how it is played.
- The entire introduction will be played for the students to understand the game dynamics.

2) Duration of the 10-hour game

- Session 1: A small stable population will be initiated.
- Session 2: Connections with neighboring towns will be developed.

- Session 3: Investments will be made in education, health, and security.
- Session 4: A railroad or highway will be constructed.
- Session 5: Construction will take place in the other neighboring cities.
- Session 6: New transportation methods will be built based on the city.
- Session 7: A balance will be sought in six aspects: traffic, health, education, environment, security, and land value.
- Session 8: Demographic growth will be initiated by expanding more areas.
- Session 9: Community problems will be addressed and solved.
- Session 10: More recreational sites and leisure areas will be created if there is stability.

The sessions are for reference purposes, as the students have the final decision on what they choose to do.

3) Objectives of the game in SimCity:

- Achieve a population of at least 50,000 inhabitants.
- Avoid having a deficit.
- Maintain a positive score of over 60% in the six aspects (traffic, health, education, environment, security, land value).

III. RESULTS

Table III shows the total scores of the pretest and posttest of Watson Glaser applied to the 25 students.

TABLE III. COMPARISON OF WATSON GLASER TEST SCORES

Student	Pretest	Posttest
01	35	52
02	33	50
03	37	51
04	32	51
05	36	50
06	36	52
07	31	49
08	33	52
09	37	52
10	37	48
11	36	54
12	36	49
13	37	50
14	36	50
15	30	46
16	34	54
17	36	54
18	30	51
19	33	52
20	34	48
21	32	48
22	32	49
23	35	48
24	32	50
25	32	48

Table IV shows the results by dimensions of the pretest Watson Glaser.

TABLE IV. RESULTS BY DIMENSIONS OF THE WATSON GLASER PRETEST

Students	Inferences	Assumptions	Deductive Reasoning	Interpretation of Information	Evaluation of Arguments	Total
01	8	8	6	6	7	35
02	6	10	5	6	6	33
03	7	8	6	7	9	37
04	5	12	6	6	8	37
05	7	9	4	6	10	36
06	7	8	5	6	10	36
07	6	7	5	6	7	31
08	7	7	4	7	8	33
09	6	8	4	8	11	37
10	6	9	5	8	9	37
11	5	9	6	8	8	36
12	7	9	5	7	8	36
13	8	10	5	7	7	37
14	8	9	5	7	7	36
15	5	7	6	5	7	30
16	8	9	6	5	6	34
17	6	11	5	7	7	36
18	6	6	5	7	6	34
19	8	7	6	6	6	33
20	6	9	5	6	8	34
21	7	8	4	6	7	32
22	5	8	6	5	8	32
23	8	11	3	7	6	35
24	6	9	3	6	8	32
25	8	5	4	6	9	32
Total	15	16	9	12	15	67

TABLE V. RESULTS BY DIMENSIONS OF WATSON GLASER POSTEST

Students	Inferences	Assumptions	Deductive Reasoning	Interpretation of Information	Evaluation of Arguments	Total
01	12	12	7	9	12	52
02	11	11	7	9	12	50
03	11	13	7	9	11	51
04	10	14	8	8	11	51
05	11	12	6	8	13	50
06	12	12	7	8	13	52
07	13	11	6	8	11	49
08	13	11	7	9	12	52
09	11	10	7	11	13	52
10	9	10	7	9	13	48
11	12	13	8	11	10	54
12	9	11	8	9	12	49
13	13	12	6	9	10	50
14	12	11	6	9	12	50
15	9	9	8	8	12	46
16	15	12	8	8	11	54
17	15	13	8	8	10	54
18	12	11	7	10	11	51
19	10	13	8	10	11	52
20	11	10	6	8	13	48
21	9	12	7	8	12	48
22	9	11	7	11	11	49
23	9	11	7	8	13	48
24	11	12	6	8	13	50
25	12	9	6	10	11	48
Total	15	16	9	12	15	67

Table V presents the results of the posttest Watson Glaser assessment, indicating a comprehensive overview of students' critical thinking skills across different dimensions. The provided scores suggest a notable improvement in various critical thinking dimensions following the implementation of the SimCity game. It is evident that the use of SimCity has positively impacted students' critical thinking abilities. Specifically, the dimensions of Inferences, Assumptions, Interpretation of Information, and Evaluation of Arguments display improved scores, underscoring the game's influence on enhancing these critical thinking aspects. Moreover, the overall total score demonstrates a collective improvement, reinforcing the notion that the game has contributed significantly to the students' overall critical thinking skills.

TABLE VI. PAIRED T-TEST

Statistics	Variable 1	Variable 2
Mean	34.28	50.32
Variance	5.376666667	4.476666667
Observations	25	25
Pearson correlation coefficient	0.363155411	-
Hypothesized difference of means	0	-
Degrees of freedom	24	-
t statistic	-31.97783414	-
P(T<=) one-tail	1.7226E-21	-
Critical value of t (one-tail)	1.71088208	-
P(T<=) two-tail	3.4452E-21	-
Critical value of t (two-tail)	2.063898562	-

Table VI displays the paired samples t-test, which shows a significant difference between both variables. This indicates that the impact of the SimCity video game application influences the development of critical thinking. It can be concluded that variable 2 has a significantly higher mean than variable 1. This conclusion is based on the extremely low p-values, the comparison of the t-statistic with critical values, and the direction of the observed difference.

1. Data Analysis

An analysis of the data collected is presented and analyzed in the following graphs.

According to Fig. 1, the development of the inference ability varied among each student, due to the multiple strategies that can be approached in the game. It is evident that the development in the inference ability of each student was positive, as all of them obtained a higher score compared to the pretest, surpassing their previous score by an average of five points. For a successful experience in SimCity, it is necessary to infer that, thanks to this ability, we can determine or identify certain information that is not explicitly stated in the source [24]. Thus, to achieve the proposed objectives in SimCity, some students approached the inference aspect better, for example, constructing multiple small and stable cities to fulfill the goals without encountering problems. Overall, the inference ability was developed through the game's challenges, which were tackled with varied solutions proposed by the players.

In Fig. 2, a lesser progress is observed in the development of the assumption ability, a factor related to how students approached problems based on clues and how effective their solutions were. Considering something as true or real based on the clues available [25]. This understanding allows us to infer those problems such as the need for more schools, shopping centers, or other facilities placed in the necessary quantity and the right location were actions taken by the students to fulfill the proposed objectives. In conclusion, the assumption ability had a more complex development, as the numerous options provided by SimCity can lead to making many errors.

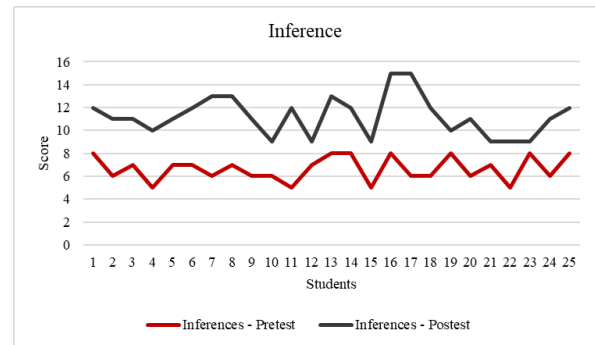


Fig. 1. Scores of the inference ability.

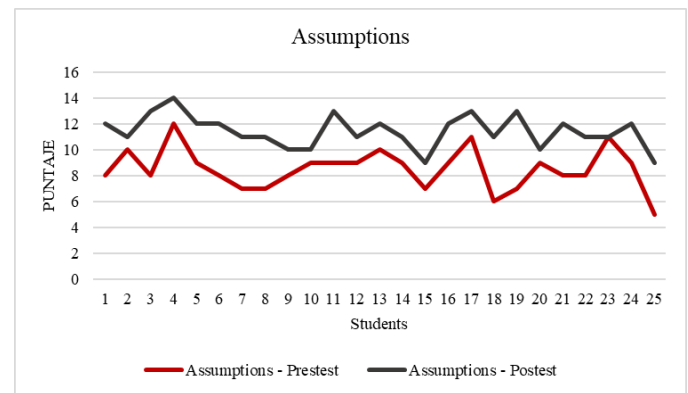


Fig. 2. Scores of the assumption ability.

According to Fig. 3, regarding deductive reasoning, a positive impact is observed as all students obtained better scores, improving by an average of 2 points compared to the pretest. Furthermore, with a maximum score of 9, significant development can be identified, with six students achieving a score of 8. In relation to SimCity, this ability is of utmost importance because when there are problems with city organization or services, the game communicates through messages that inform the player about possible strikes or riots, which leads students to make decisions and infer conclusions about the potential outcomes if the problems are not resolved. This reasoning allows organizing premises into syllogisms that provide decisive proof for the validity of a conclusion; it is often said to deduce from an unexplained situation [26]. For example, if there are environmental problems, the premises can be the notifications and the low percentages in city aspects such as health and the environment. Overall, SimCity enables the exercise of deductive reasoning through its notification system and the overall panel displaying the city's aspects.

According to Fig. 4, in terms of information interpretation, the progress was similar to deductive reasoning, with an average increase of 2.48 points from the previous score. This ability is related to the concepts provided at the beginning of the game application since the interpretation of the given tutorial information, which aimed to provide foundations on how a city can grow sustainably, also offered solutions to common problems. The small difference in scores obtained by the students demonstrates a similar interpretation of the tutorial. However, SimCity is not a game of narratives or closed decision-making; it is an open-ended game. Therefore, the interpretation of information varied in each case, as the developing city differentiated itself from others. The information provided to the students through notifications was diverse, and thus, its interpretation also varied.

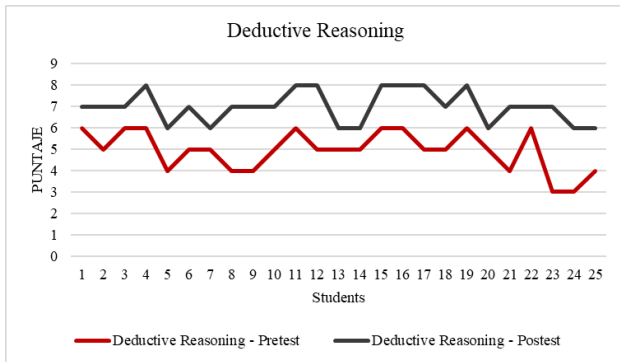


Fig. 3. Scores of the deductive reasoning ability.

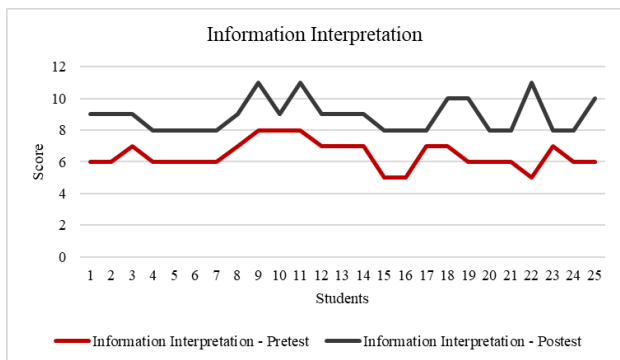


Fig. 4. Scores of the information interpretation ability.

Fig. 5 shows an average development of 4 points compared to the pretest, which indicates that despite not being a game solely focused on answering questions based on information, SimCity has demonstrated that errors allow for significant student learning. An argument is a reasoning that justifies something, and in the case of SimCity, when a decision deviates from the intended path, it can be considered as something negative, leading to the evaluation of the student's decisions. This process prompts students to evaluate their own arguments and make decisions based on them. This attribute is present in each student during the game, as problems frequently arise in the process of building a stable city. Therefore, students often seek solutions, and if they make an incorrect decision, they realize it through the economic deficits they encounter, creating the need to evaluate their decisions

and find the best solution, which students achieved and is evidenced by the improvement in their scores.

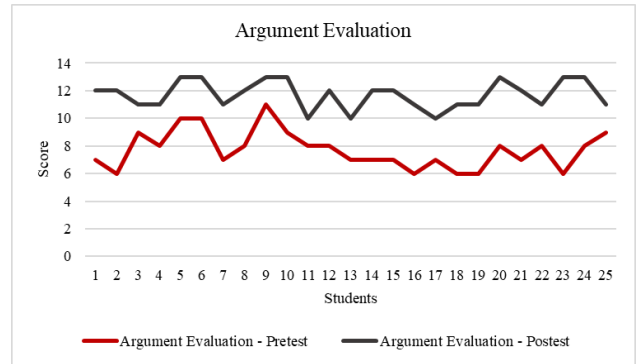


Fig. 5. Scores of the argument evaluation ability.

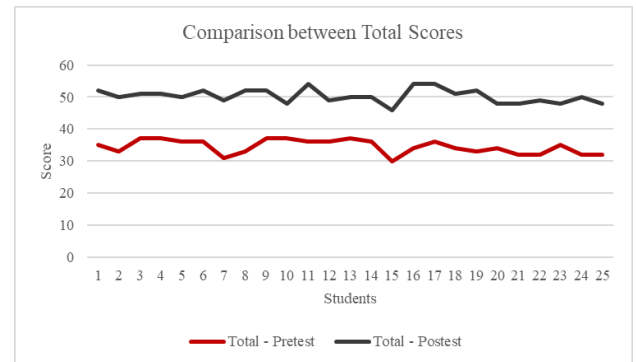


Fig. 6. Comparison of the total scores between the pretest and posttest.

According to Fig. 6, it is evident that critical thinking has developed in the students, as all of them showed improved scores. These results infer that the SimCity video game had a positive impact on the development of critical thinking in students.

IV. DISCUSSION

The results of this study are in line with previous research highlighting the positive impact of the game SimCity on the development of specific cognitive skills. According to the work done by [20], a significant effect of the SimCity game on the enhancement of students' spatial intelligence is established. This finding is important as it supports the notion that video games can contribute to the development of specific cognitive skills, in this case, spatial intelligence, which is crucial for effectively understanding and navigating spatial environments.

Furthermore, the results obtained in this research align with the findings of [21], who also emphasized the perception of SimCity as a motivating and enriching tool in the educational scenario. The motivating diversification of learning adds to the evidence that video games, in this case, SimCity, can generate greater student engagement and participation in the teaching-learning process, potentially improving information retention and the understanding of complex concepts.

Regarding our findings, it was found that the use of the SimCity video game in teaching-learning processes has a positive impact on the development of critical thinking skills in students of regular basic education, although it is partial. This

suggests that while the game contributes to the overall development of critical thinking, some dimensions of critical thinking may not be comprehensively addressed. The dimensions of critical thinking are varied and may require specific approaches for complete development.

In particular, the logical dimension showed the greatest development through the use of the SimCity video game, and this dimension exhibited a close relationship with the results obtained in the Watson Glaser test. This correlation highlights the importance of logical thinking in problem-solving and the evaluation of arguments, two central aspects of critical thinking. However, it is observed that other dimensions of critical thinking could benefit from complementary pedagogical approaches to achieve comprehensive development.

According to [27], engaging in the SimCity video game activity has provided a framework for developing meaningful educational situations that facilitate the contextualization and application of scientific conceptual content acquired by students. Additionally, according to [28], the development of critical thinking attitudes can be determined through various aspects of the gaming experience. One of these aspects involves analyzing the underlying model in the video game and comparing it with reality, while the other aspect involves exploring the values and counter-values presented. In this way, students have applied their acquired knowledge to compare simulation and reality, noting significant differences. Some of the weaknesses criticized for potential educational use can thus be transformed into opportunities for developing skills and attitudes.

For example, it is criticized that the model programmed in the simulation never reaches reality [29], [30]. The development of problem-solving skills through the SimCity video game series has already been confirmed in studies by [31] and [32]. In this case, the steps that students take to solve problems of varying difficulty, ranging from easy to medium and complex, have been analyzed to determine the process followed, in addition to the final result. The conclusion is that through the simulation video game, students can employ very different strategies to reach the same final solution, thus fostering creativity and accommodating the diversity of students [33]. As indicated by [34] and [35], an added value for improving the teaching-learning process of Geography is provided by the need to manage a simple layered geographic information system, which can serve as an introduction to GIS (geographic information systems). Furthermore, according to [36], [37] the development of problem-solving skills is one of the main consequences of becoming a mayor to manage and build a city with SimCity. The research conducted in contrast to the reviewed background allows us to glimpse that it is possible to use this type of video game to develop critical thinking skills in students, as demonstrated by the results.

V. CONCLUSIONS

These collective results underscore the effectiveness of the SimCity game in fostering critical thinking skills among students. The improvements across dimensions and the significant shift in mean scores support the premise that video games, when strategically integrated into educational contexts,

can indeed contribute to the development of targeted cognitive abilities. This study adds valuable insights to the growing body of research on the educational potential of video games and suggests that their incorporation into pedagogical approaches warrants further exploration and consideration.

The use of the SimCity video game in teaching and learning processes allows for the partial development of critical thinking skills in regular basic education students since the focus is not on each dimension of critical thinking as such, given their diversity. Similarly, the logical dimension was the most developed through the SimCity video game, being the one most related to the Watson Glaser test.

It is inferred that with adequate stimulation through the SimCity video game, critical thinking can have a moderate and effective development in regular basic education students, as significant progress in scores was observed based on the data obtained from the post-test, which were validated with the hypothesis through the Student's t-test.

Likewise, the effectiveness of SimCity is more reflected in inferences and argument evaluations, as greater progress was observed during the post-test evaluations compared to other abilities, even resulting in ideal scores for two students in the inference capacity.

On the other hand, the progress in information interpretation was lower compared to other abilities, as it had a lower but achieved progress. This is because the video game initially only presents a tutorial, which provides a broad scale understanding of the theme, resulting in minimal information to interpret. Meanwhile, the use of skills such as deductive reasoning, inferences, and argument evaluation is more necessary for the development of critical thinking, as only the best decisions ensure the achievement of the proposed objectives.

RECOMMENDATIONS

Regarding the SimCity video game, it should be played in a supervised environment to achieve better results, as this ensures the proper use of the video game since, despite not being an online game, uncontrolled use can lead to addictive behavior. Additionally, it is recommended for future research to apply the test to a control group that does not play SimCity to enable a comparison with an experimental group.

The research was conducted with a relatively small sample of students, which could limit the generalization of the results to larger populations.

RESEARCH LIMITATIONS

Limited Duration: The study was carried out over a specific period of time, which might not fully reflect the long-term effects of implementing SimCity on students' critical thinking.

Specific Context: The results could be influenced by the specific conditions and characteristics of the educational institution where the study was conducted, making it challenging to extrapolate the findings to other settings.

Focus on a Single Tool: The research centered on the SimCity game as the sole intervention tool. Including other

tools or pedagogical approaches could have provided a more comprehensive understanding of how critical thinking skills are developed.

Impact of External Factors: Uncontrolled external factors, such as individual student motivation or influences beyond the educational environment, could have influenced the outcomes.

Measurement of Other Variables: The research specifically focused on critical thinking and did not consider the measurement of other skills or competencies that could have been influenced by the game.

REFERENCES

- [1] J. Vives, «Serious Games, juegos con finalidad educativa y terapéutica,» 2018. <https://bit.ly/3rZzKfV>
- [2] Z. Tanes y Z. Cemalcilar, «Learning from SimCity: An empirical study of Turkish adolescents». 2010. <https://doi.org/10.1016/j.adolescence.2009.10.007>.
- [3] T. Friedman, «The Semiotics of SimCity». 1999. <https://bit.ly/3OI9Tlw>
- [4] K. Salen y E. Zimmerman, «Rules of play». 2004. <https://bit.ly/3KmoII4>
- [5] M. Lauwaert, «Challenge Everything? Construction Play in Will Wright's SimCity». 2007. <https://doi.org/10.1177%2F1555412007306205>.
- [6] «SimCity 4,» 2017. <https://bit.ly/3rZAmSL>
- [7] M. Romero y O. Turpo Gebera, «Serious Games para el desarrollo de las competencias del siglo XXI». Revista de Educación a Distancia. 2015. <https://revistas.um.es/red/article/view/233511>
- [8] Á. V. Jusino, «Teoría y pedagogía del pensamiento crítico». 2013. <http://pepsic.bvsalud.org/pdf/pp/v3-4/v3-4a04.pdf>.
- [9] A. Triglia, «Las 4 etapas del desarrollo cognitivo de Jean Piaget,». 2015. <https://bit.ly/3OI4P0r>
- [10] Y. P. Zapata Maya, «La formación del pensamiento crítico: entre lipman y Vygotski,» 15 febrero 2010. <https://bit.ly/3QsaLMA>.
- [11] D. R. Paul y L. E. Elder, «La mini-guía para el Pensamiento crítico Conceptos y herramientas». 2013. <https://bit.ly/45fuWRI>.
- [12] E. Toro Toloza, R. Ponce Alvarado, R. Ramírez Castro y J. Navia Alava, «Pensamiento crítico-complejo-innovador: reencuentro con una nueva pedagogía». 2019. <https://doi.org/10.46377/dilemas.v30i1.1175>
- [13] C. Rojas Osorio, «¿Qué es pensamiento crítico? sus dimensiones y fundamentos histórico-filosófico?» 2013. <https://www.redalyc.org/pdf/1942/194220390001.pdf>
- [14] I. E. Gámez, «La aplicación didáctica de "SimCity 4" en la formación universitaria: el caso de la Facultad de Administración de la UV, región Veracruz». 2010. <https://bit.ly/3QpQTcU>
- [15] HQ Autor Psychology Notes, «Las etapas de Piaget del desarrollo cognitivo». 2018. <https://www.psychologynoteshq.com/piagetstheory/>
- [16] E. Arteaga y V. Torres, «Videojuegos y habilidades del pensamiento». RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo 2018. <https://doi.org/10.23913/ride.v8i16.341>
- [17] D. Lobo, «La ciudad no es un juguete,» 2006. <https://dialnet.unirioja.es/>
- [18] P. L. Gutiérrez, «Aplicación de inteligencia artificial en videojuegos». 2017. <https://bit.ly/3ORHny9>
- [19] J. Esteban Rodríguez, «Los videojuegos de estrategia como herramienta para el desarrollo de competencias en la toma de decisiones». 2018. <https://bit.ly/3QIXLYN>
- [20] E. Putra, A. T. Bima and R. Mamat. "The Effect of Simcity as Instructional Media in Geography Learning on Learners' Spatial Intelligence." Proceedings of the 2020 International Conference on Education Development and Studies. 2020. <https://10.1145/3392305.3396896>
- [21] U. Arnold, S. Heinrich and R. Maria. "SimCity in infrastructure management education." Education Sciences 9.3 (2019): 209. <https://10.3390/educsci9030209>
- [22] Z. R. Vargas-Cordero, «La investigación aplicada: una forma de conocer las realidades con evidencia». Revista Educación, vol. 33, núm. 1, pp. 155-165 Universidad de Costa Rica. 2009. <https://www.redalyc.org/pdf/440/44015082010.pdf>.
- [23] A. & J. W. Burton, «Adult norms for the Watson-Glaser Tests of Critical Thinking». The Journal of Psychology, pp. 43-48., 2015. <https://10.1080/00223980.1945.9917217>
- [24] I. Rovira-Salvador, «Pensamiento inferencial: qué es y cómo desarrollarlo». 2018. <https://psicologiymente.com/inteligencia/pensamiento-inferencial>.
- [25] Real Academia Española, «Diccionario de la lengua española, 23.ª ed,» 2019. <https://dle.rae.es>.
- [26] G. Dávila Newman, «El razonamiento inductivo y deductivo dentro del proceso investigativo en ciencias experimentales y sociales,» 2006. <https://www.redalyc.org/pdf/761/76109911.pdf>.
- [27] E. y. J. A. Nilsson, «Simulated sustainable societies: Students' reflections on creating future cities in computer games». Journal of Science Education and Technology, pp. 33-50, 2019. <https://eric.ed.gov/?id=EJ913117>
- [28] J. J. Clemente, «El videojuego SimCity como recurso para la enseñanza-aprendizaje de la Geografía en Bachillerato,» de XIX Congreso Internacional de Tecnologías para la Educación y el Conocimiento, Madrid, 2019. <https://10.13140/RG.2.1.2901.9042>
- [29] L. Daniel, «La Ciudad no es un juguete, cómo SimCity Juega con el urbanismo». Arquitectos: información del Consejo Superior de los Colegios de Arquitectos de España, pp. 59-66., 2018. https://www.daquellamanera.org/files/Lobo_SimCityCiudadJuguete06.pdf
- [30] S. y. T. M. Rufat, «Jeux vidéo et simulations urbaines: trucs ou astuces?». Cybergeo. Revue européenne de géographie, pp. 50-62, 2014.
- [31] P. Adams, «Teaching and learning with SimCity 2000». National Council for geographic education., pp. 47-66, 1998.
- [32] Y. Carolyn Yang, «Building virtual cities, inspiring intelligent citizens, » Digital games for developing students' problem solving and learning motivation, pp. Computers & Education, n° 59. 365-377 Páginas. Edita: Elsevier, 2018.
- [33] N. M. L. y. L. P. Monjelat, «Paso a paso: aprendiendo a resolver problemas con SimCity Creator,» Actas del II Congreso Internacional de videojuegos y educación., pp. 178-200, 2014.
- [34] H. y. L. Y. Lin, «Digital educational value hierarchy from a learners' perspective». Computers in Human Behavior, pp. 1-12, 2019.
- [35] M. Monguillot Hernando, C. González Arévalo, C. Zurita Mon y Almirall, «Play the Game: gamificación y hábitos saludables». 2015. <https://www.redalyc.org/pdf/5516/551656902003.pdf>.
- [36] K. Terzano y V. Morckel, «SimCity in the Community Planning Classroom: Effects on Student Knowledge, Interests, and Perceptions of the Discipline of Planning». 2016. <https://doi.org/10.1177/0739456X16628959>
- [37] B. Maraza-Quispe et al., "Towards the development of research skills of physics students through the use of simulators: A case study," Int. J. Inf. Educ. Technol., vol. 13, no. 7, pp. 1062-1069, 2023. <https://10.18178/ijet.2023.13.7.1905>

An Automated Medical Image Segmentation Framework using Deep Learning and Variational Autoencoders with Conditional Neural Networks

Dustakar Surendra Rao¹, L. Koteswara Rao^{2*}, Bhagyaraju Vipparthi³

Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation,
Hyderabad, 500075, Telangana, India^{1,2}

Department of Electronics and Communication Engineering, Guru Nanak Institutions Technical Campus,
Hyderabad, 501506, Telangana, India¹

Department of Electronics and Communication Engineering, Siddhartha Institute of Engineering and Technology,
Hyderabad, 501506, Telangana, India³

Abstract—It is a highly difficult challenge to achieve correlation between images by reliable image authentication and this is essential for numerous therapeutic activities like combining images, creating tissue atlases and tracking the development of the tumors. The separation of healthcare data utilizing deep learning variational autoencoders and conditional neural networks is presented in this research as a paradigm. One of the essential jobs in machine vision is the partitioning of an image. Due to the requirement for low-level spatial data, this assignment is more challenging compared to other vision-related challenges. By utilizing VAEs' capacity to develop hidden representations and combining CNNs in a conditioned environment, the algorithm generates accurate and efficient results for the segmentation. Moreover, to learn the representation of latent space from labelled clinical images, the VAE is trained as part of the system that is suggested. After that, the representations that were learned and real categorizations are used to develop the conditional neural network. Furthermore, the model that has been trained is utilized to accurately separate the areas that are important in brand-new medical images during the inferential stage. Thus, the experimental findings on several healthcare imaging databases show the enhanced precision of segmentation of the structure, highlighting its ability to enhance automated diagnosis and treatment. Henceforth, the suggested Deep Learning and Variational Auto Encoders with Conditional Neural Networks (DL-VAE-CNN) are employed to solve the pixel-level problem of classification that plagues the earlier investigations.

Keywords—Deep learning; variational autoencoders; CNN; medical image segmentation; automated diagnosis and treatment

I. INTRODUCTION

A harmful disease that has one of the highest five-year rates of survival is brain tumor. To determine the kind of brain tumor, neurological specialists frequently employ magnetic resonance imaging (MRI). Despite the amazing advancements in science over the past few years, certain diseases still pose a grave risk to life [1]. Of these, cancer of the brain is among the most aggressive. Brain tumors are the result of untamed, erratic polypeptide development inside and outside the brain's cells. Carcinoma brain tumors are the most dangerous kind, but benign tumors can also exist. Thus, brain cancer is the

common name for the malignant variety of cerebral tumor. A tumor is deemed cancerous if it emerges from the membrane and invades nearby tissues. There are three main types of brain tumors. They are gliomas, pituitary tumors and meningiomas. The pituitary brain tumor is a malignant development of polypeptide that surrounds the pituitary gland, which is a gland that is situated at the bottom of the brain. On the outermost layer of the brain below the skull, meningioma, a benign tumor, is detected. It grows slowly. Amongst all brain tumors, glioma has the greatest fatality rate globally. Pituitary and meningioma tumors can be easily found due to the position of the different types of brain tumors; however, gliomas are challenging to find and study.

The objective of a semantic segmentation is to define the limits of important anatomical or pathological features [2]. The findings can be applied to multisensory image authorization, therapeutic targeting preparation, or operational systems for navigation. Furthermore, the field of imaging in medicine has made extensive use of its versions for segmentation. Due to the increasing adoption of Computed Tomography (CT) and Magnetic Resonance (MR) imaging for medical diagnosis, treatment planning and medical inquiries, image segmentation is necessary for radiological experts to effectively utilize technological advances to assist in helpful assessment as well as the planning of treatment. Delineating anatomical features and other areas of interest (ROI) requires the use of trustworthy methods [3]. Moreover, the utilization of three-dimensional (3D) convolutional neural networks for tissue separation on CT data is very efficient: however the process of creating labels for training is time-consuming. Therefore, the effectiveness of the algorithm that has been trained is hampered by a lack of designations, particularly when it comes to its ability to generalize to everyday clinical practice. Nevertheless, due to the complexity of person's structure and pathological situations, numerous variations and medical conditions are not included in this small training information and that may partially account for the reported discrepancy among efficiency utilizing real-world information and anticipated results [4]. Even if trained on a greater amount of data sets, there will be discrepancies among the model's probability and the real-world probability. These changes

could relate to technological issues, individual traits or more uncommon pathological diseases.

Nevertheless, pathological identification is vital yet a challenging task in the computation and evaluation of clinical images. Although certain tasks such as tumor measurements or therapy surveillance refers to precise segmentation of the conditions and other applications such as assisting the physician with inspecting the image or retrieving visuals with diseases at particular points from a data base and it only require information to identify the illness and an approximate representation of its description [5]. Moreover, researchers concentrate on the identification of images with diseases as another instance of such an application. Furthermore, missing connections among images due to pathological features could result in incorrect identification that is leading to the surrounding regions. In order to estimate the regions of lacking connections, it might be good to incorporate the earlier pathological data into the process of registration. Therefore, variational autoencoders (VAEs) are broad dormant area computational simulations that have proven incredibly effective in a variety of fascinating applications in medical information technology including large-scale physiological order evaluations, incorporated multi-omics data analysis and the design of molecules and protein layout [6].

Additionally, the primary concept behind VAEs is to comprehend the data distributions in a manner that enables the generation of fresh, useful information with higher intra-class variability from the transmitted population. Moreover, deep generative algorithms have received an abundance of interest previously as a result of their numerous information creation capabilities. Therefore, one of the most well-liked methods of dynamic modelling and low-dimensional network representational learning amongst them is the variational autoencoder (VAE). Nowadays, diagnostic imaging is commonly employed to locate tumors and other anomalies in the body of a person for investigation and research. According to the illness type and the patient's body portion, several techniques for medical imaging are employed for diagnostics [7]. To diagnose an anomaly promptly and accurately, scanning needs to be done. The procedures include computer tomography (CT) that is frequently utilized to compare numerous soft tissue types including the liver, the lung cells and lipids, as well as X-ray, that is utilized for recognizing fractures in bones or to identify tissues that are hard. Positron emission tomography (PET) scans and MRI scans are both frequently employed in the recognition of anomalies in the fields of neuroscience, heart disease, malignancy and connective tissue impairment respectively.

Henceforth, a convolution's job is to find structures in the data it receives that represent characteristics that can be detected. Convolutional operations require kernels, also known as filtration systems. A convolution execution is viewed mathematically and the outcome of multiplying an element-by-element portion of the input by the filter adds the outcomes. Until the entire input area has been covered, the procedure continues by moving the location of the fragment as determined by the pace a number of instances. This creates a new result array by computing just one value for every

segment [8]. Furthermore, with pooling, linguistically associated characteristics are combined into one, as opposed to convolutions. In order to decrease the size of the module and create constancy to minor changes and deformations, they are termed as the fixed processes which are not trainable. Hence, CNNs have shown they can produce images of outstanding quality with significantly fewer training time and computing demand. However, fully connected networks can also produce images. Based on vector data an image is used as input, there are two distinct kinds of picture generating algorithms. Though, improvements in technology have made it easier to acquire an image, which has resulted in the production of enormous numbers of pictures with excellent resolution at extremely low costs. The creation of biological algorithms for image processing has significantly improved as a result of this and it has made it possible to create computerized methods for gathering data through visual inspection or assessment [9].

The key contribution of the established outline is updated as follows:

- At first, the data has been collected from the dataset.
- Next, the pre-processing step is completed by Histogram Equalization.
- After that, the segmentation of the image is done by K means clustering.
- Feature extraction process is done by VAE, this will model the data.
- Classification is done by Convolutional Neural Network.
- Finally, the effectiveness of the proposed approach has been acknowledged and compared with various methods to demonstrate its superiority in efficiency and productivity.

The other parts of this research are broken down into the following divisions as an outcome: The related works are revealed in Section II after an in-depth examination. Problem statement is covered in Section III. In Section IV, the specifics of DL-VAE-CNN are examined. The experiment's results are analyzed and reported in detail in Section V. The research's conclusion is found in Section VI.

II. RELATED WORKS

Chen et al.[10] disclosed in his paper that, the development of novel healthcare image processing techniques has attracted a lot of academic curiosity in deep learning, and this deep learning-based algorithm have achieved amazingly well across a range of healthcare scanning applications to enhance illness identification and detection. Considering their achievement, the absence of sizable and thoroughly structured data sets severely hinders the advancement of deep learning algorithms for medical image interpretation. In order to give an in-depth account of using deep learning techniques in diverse medical imaging analysis assignments, recent results are utilized in this study. Moreover, researchers focus in particular on the most recent developments and contributions made by state-of-the-art unsupervised and semi-supervised

deep learning in the analysis of healthcare images that are outlined according to various application circumstances, encompassing categorization, the process of segmentation identification and registration of images. The inaccessibility problem is the research's main flaw.

Considering pixel-based deformation for evaluation, there is currently an exciting opportunity on super-resolution images with bigger up-scaling. This is due to the significant improvement in empirical precision, said by [11]. In addition to the very soften impact, the perceived resemblance is poorly grasped. Because of the development of generative adversarial networks, it is now feasible to super-resolve low-resolution images to produce distribution-sharing photo-realistic images. Generative networks, nevertheless, struggle with mode-collapse issues and unsustainable sampling creation. Owing to the probabilistic dispersion of the high-resolution images produced by the low-resolution images, researchers suggest doing Image Super Resolution via Variational Autoencoders (SR-VAE) learning. Moreover, researchers incorporate the conditioned sampling method to reduce the implicit substructure for rebuilding since Conditional Variational Autoencoders frequently produce blurry images. Thus, researchers estimate the difference among hidden vectors and the conventional Gaussian distribution using KL damage to assess the model's generalization. Finally, with the goal to negotiate the trade-off among perceptions and super-resolution deformation, researchers calculate the reconstructed image using both the revised deeper component reduction among SR and HR images as well as pixel-based reduction.

Uzunova, Ehrhardt, et al. [12] revealed in his paper that, the capacity to decode black box artificial intelligence approaches for analyzing medical images is becoming increasingly important. A system like trained neural network must be prepared to justify its choices and projections in order to be trusted by a physician. In this study, researchers take on the challenge of coming up with logical justifications for the conclusions reached by healthcare image classification algorithms, which are trained to distinguish among various illnesses and tissues that are healthy. Finding out if the results of classification alter when such image areas are removed is a logical way to figure out the fact that areas of the image have an impact on the training classification. This concept can be effectively put into practice if it is expressed as a reduction issue. This is considered as the drawback here. In order to make a difference, researchers define the omission of pathologies as the substitution of those conditions with a substitute produced by variational autoencoders that appears normal. The investigations using a classification neural network on brain lesion MRIs and OCT (Optical Coherence Tomography) images demonstrate that an important substitution of "removed" image areas has a substantial influence on the accuracy of the given interpretations. Thus, the suggested omission procedure has been demonstrated to be effective when contrasted with four other tried-and-true techniques, this approach yields superior outcomes.

Higher dimension hyperspectral images always need more computing, which renders the processing of images difficult said by [13]. Deep learning methods have excelled in many areas of processing images and they are useful for enhancing

the accuracy of classification. The complete extraction of copious spectrum data, including the fusion of geographical and spectral information, still presents significant problems. This research proposes an innovative design for autonomous hyperspectral extraction of features that utilizes the spatially reviewing variation auto encoder (AE). The disadvantage in this situation is that having numerous channels that makes it hard to recognize materials and continually wastes resources. This technique's main idea is to refine the acquired spectral characteristics by collecting spatial characteristics from multiple facets using created systems. A multilayered generator collects spectral characteristics based on the data and standard errors it produces, space-time vectors are created. With the ability to modify the derived mean vectors, multilayered convolutional neural networks and long short-term memory systems are used to gather spatial information utilizing local perception and consecutive perception. Additionally, the suggested function of loss ensures the coherence of the likelihood probabilities of different implicit spatial characteristics that were acquired from the exact same neighbor area. The effectiveness of this approach is enhanced by the incorporation of spatial extraction of features algorithms and deep AE designs, which are built specifically for hyperspectral images.

Peyré and Cuturi [14] disclosed in his paper that, the generative versus discriminatory modelling is a key distinction in the arena of machine learning. In contrast to generative modelling that seeks to tackle the broader issue of developing an aggregate allocation over all the variables, discriminatory modelling strives to train an indicator provided the data available. A model that generates information imitates how data is produced in reality. Moreover, a sound architecture to develop profound latent-variable systems and related interpretation frameworks is provided by variational autoencoders. In this paper, researchers introduce variational autoencoders and discuss several significant improvements.

With multiple uses including scenario comprehension, medical imaging analysis, robotics awareness, surveillance footage, virtual reality and image compression as well, segmenting images is a fundamental subject in the fields of image processing and vision for computers said by [15]. In the fiction, a total of image separation techniques have been generated. There has been a substantial quantity of determinations subsequently focused on improving algorithms for image segmentation using models trained with deep learning as a consequence of the efficacy of deep learning algorithms in an array of visual application. Using a wide range of groundbreaking research on conceptual and instance-level classification such as entirely convolutional pixel-labeling systems, encoder-decoder designs, multi-scale and pyramid-based methods, intermittent systems, graphic focus designs and more researchers offer an in-depth examination of the research available as of the time of publication in this analysis. Moreover, researchers evaluate the similarities, advantages, and disadvantages of different deep learning simulations, look at the most popular data sets, present results, and talk about possible future study avenues in this field. The classification issue is the research's main flaw.

Wang et al. [16] said in his paper that, the healthcare segmentation of images has made widespread usage of deep learning, and many studies have been distributed authenticating the knowledge's presentation in the arena. Deep learning algorithms for medical image segmentation are described in a thorough topical analysis. This essay offers two distinctive contributions. Researchers categorize the most recent literary works according to a multi-level architecture from rough to satisfactory, in contrast to previous studies that explicitly split studies of deep learning on segmenting medical images into multiple categories and present studies thoroughly for each category. Secondly, while unstructured methods have been covered throughout numerous previous surveys and aren't currently in demand, this work concentrates on supervised and poorly supervised learning methods. Also, researchers examine research on supervised learning methods in three areas: the choice of backbone networks, the layout of network wedges, and the enhancement of function losses. Researchers go into the research on poorly supervised learning methods independently for data enhancement, learning by transfer, and dynamic categorization. This study organizes the literatures quite differently from previous investigations. It is also easier for users to comprehend the pertinent justification, which will help them come up with suitable deep learning-based advances for healthcare image segmentation.

III. PROBLEM STATEMENT

Correct interpretation in remedial and diagnosis activities depends on an accurate MRI. The precision of deep learning (DL) rebuilding, is mostly unknown due to underestimating after MRI recording and the over parameterized and opaque characteristics of DL [17]. This work intends to measure the degree of ambiguity in DL model-based image restoration. The present investigation employs an automated medical

image segmentation framework using deep learning and variational autoencoders with conditional neural networks to address this uncertainty problem.

IV. METHODOLOGY

The suggested method is projected in Fig. 1. An automated medical image segmentation framework using Deep Learning and variational autoencoders with conditional neural networks (DL-VAE-CNN) is employed for this method. Regarding the purposes of training and testing, a variety of databases are accessible. To investigate the data and establish accurate labeling in this situation, the Low-Grade Gliomas (LGG) segmentation dataset is used. Furthermore, pre-processing procedures are used to increase the image precision and quality. Fig. 1 shows the proposed DL-VAE-CNN architecture.

A. Data Collection

The data set employed for the testing and training purpose is Low Grade Glioma (LGG) segmentation dataset and the data are obtained from The Cancer Imaging Archive (TCIA) brain MR images. Approximately 120 patients' brain image data is second-hand in this investigation. From that image 60% of images were designated for training process and 60% of images were designated for testing process [18]. It embraces gray scale images.

B. Pre-Processing

Histogram Equalization (HE) is used to improve the quality of images. This is done by levelling, the grey levels of the image pixels so as to continuously rearrange them within the location [19]. The image that was entered as input is turned into a final image after the histogram has been analyzed and normalized for sum computation is shown in Fig. 2.

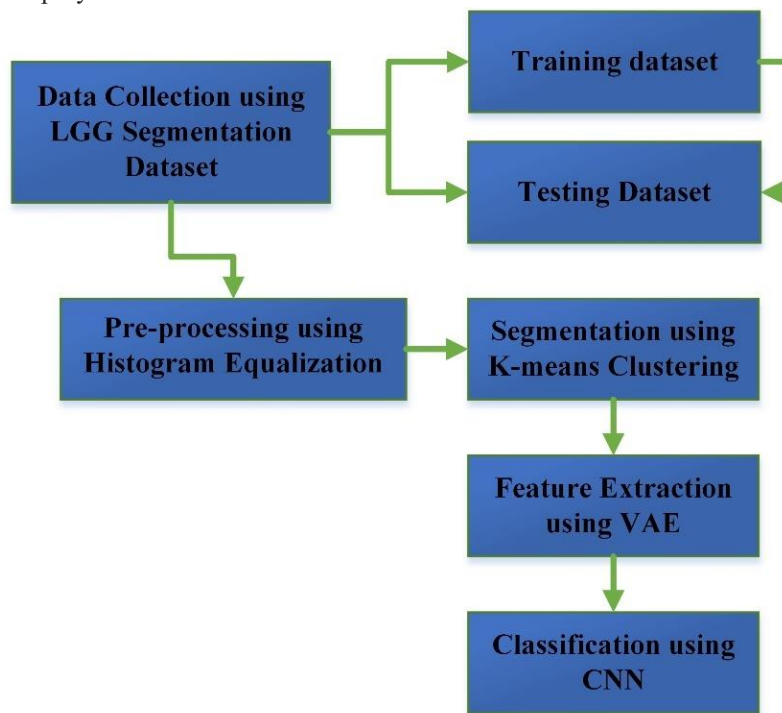


Fig. 1. Proposed DL-VAE-CNN architecture.

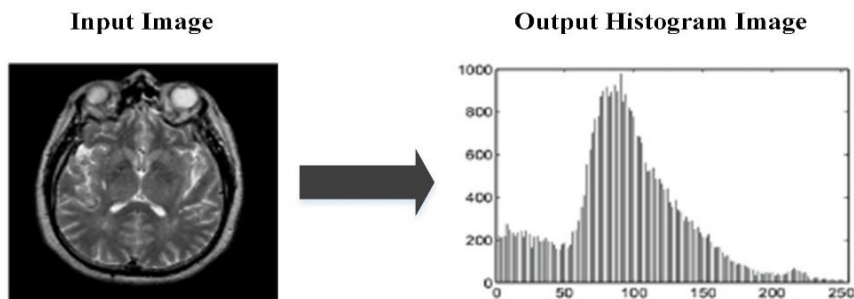


Fig. 2. Histogram equalization for healthcare image.

The method for altering the input image into output image using HE is given in the Eq. (1) below:

$$\text{Histogram Equalization image} = \text{round} \left(\frac{\text{CDF of image} - \text{CDF}(\min)}{(W \times H) - \text{CDF}(\min)} \times (G - 1) \right) \quad (1)$$

Where, CDF is denoted as Cumulative Distribution Function; cdf (min) is denoted as the minimal non-zero value of Cumulative Distribution Function; W is denoted as the width of the image; H is denoted as the height of the image; G is denoted as the greyscale of the image.

C. Segmentation using K-Means Clustering

As a key step in many applications, segmentation has grown with the importance in the context of image processing including image recovery, classification, and recognizing objects [20]. After the pre-processing step is finished, the images are separated; then, the suggested method uses K-means clustering to find comparable areas, combine them and examine each segment properly. Moreover, the regions of interest in the grouped images are transformed to RGB layout and each predicted zone of interest is calculated by fusing attributes like color and intensity ratio.

The basic goal of the K-means clustering approach is to transform unknown data set facts into distinct group components. This will address issues related to data science or clustering. This clustering's primary significance is the fact that it will ensure resolution. It will ensure that the centroids' orientations that commence out smoothly. This centroids-calculating method iterates until it determines the right centroid. It will make the clusters of various sizes and forms simpler. The geographic coordinates of the image ought to try to find $i \times j$ first, after which the resulting images should be put together to create the k-clusters. Take into account that P_k is the function that emphasizes the set of data in pixel $h(i, j)$. The Euclidean distance equation is given in Eq. (2):

Step 1: It is necessary to identify the center and the cluster K.

Step 2: For each pixel, the Euclidean distance E_d is indicated and is uttered in the Eq. (2):

$$E_d = ||h(i, j) - P_k|| \quad (2)$$

Step 3: Most of the pixels are assigned to the center point using the E_d foundation.

Step 4: The work allocated to each pixel is completed, and the new center coordinates are recalculated by Eq. (3):

$$P_k = \frac{1}{k} \sum_{j \in P_k} \sum_{i \in P_k} h(i, j) \quad (3)$$

Step 5: Continue doing this until the requirement is met.

Were,

E_d is denoted as the Euclidean distance; P_k is denoted as the group of clusters; $h(i, j)$ is denoted as the information pixel.

D. Feature Extraction using Variational Autoencoder

The procedure of choosing and displaying the most important facts or trends from unprocessed data is referred to as feature extraction [21]. It entails converting the initial information into a more condensed and useful form that may be applied to activities requiring modelling or further analysis. To improve algorithmic performance and efficiency, feature extraction is frequently used in machine learning and pattern recognition applications.

A finding in the latent space can be described probabilistically using a variational autoencoder (VAE) [22]. Thus, a distribution of probability for each underlying feature is employed to create the encoder instead of one that produces only one number to characterize every latent state feature. When using VAE, instances from the identical class wind up very near one another in the coding space, allowing for improved unsupervised representation learning. Moreover, the VAE is used in this research because it is used to detect the abnormalities in the medical images. The architectural diagram of VAE is shown in Fig. 3 and its components are explained below.

1) *Input*: The purpose of the application and area specificity determine the input to a Variational Autoencoder (VAE). The input for image-based VAEs, on the other hand, is frequently made up of images or patches of images.

2) *Encoder*: The encoder converts the input facts into the dormant space parameters associated with the distribution of probabilities. It usually consists of numerous levels of neural networks that gradually reduce the degree of dimensionality of the input data, including convolutional or fully connected layers. A collection of mean and variance vector that reflect the properties of a Gaussian distribution with multiple variables in the dormant space are the encoder's outcome.

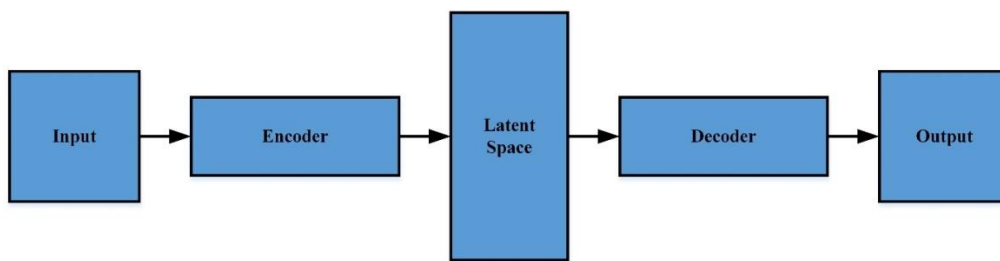


Fig. 3. Variational auto encoder architectural diagram.

3) *Latent space*: Every point in the latent space represents a latent code and is a reduced dimensional model of the input data. Moreover, the encoder gains the ability to create latent codes through training that precisely represent significant input data characteristics. Finding a clear format that includes the crucial data while eliminating the background noise is the goal.

4) *Decoder*: A sample of data from the latent space is taken by the decoder, which then transforms it to the initial input region. The decoder is prepared up of a number of strata of neural networks, similarly like the encoder and it gradually up samples the latent code to generate its result. The goal of the decoder's final product is to precisely reproduce the initial input information.

5) *Output*: A reconstruction of the incoming data is exactly the decoder delivers. The reconstruction loss, that gauges the variance among the input and the output, is the parameter that the VAE is trained to reduce. The VAE develops to produce useful reconstruction from latent codes by minimizing this loss of information.

The VAE possess some equations and is given under Eq. (4) and (5),

$$\mu_{pp}, \sigma_{pp} = R_{pp}(B_1, Y); \mu_{pv}, \sigma_{pv} = R_{pv}(B_2, Y) \quad (4)$$

$$\mu_{qq}, \sigma_{qq} = R_{qq}(B_2, Y); \mu_{qv}, \sigma_{qv} = R_{qv}(B_2, Y) \quad (5)$$

Where, R_{pp} and R_{pv} are determined as separators; B_1 and Y are the input and output latent variables; μ_{pp}, σ_{pp} are determined as the structural specific component; μ_{pv}, σ_{pv} are determined as the universal component.

6) *Loss function*: Variational Autoencoders (VAEs) train the algorithm using a loss function [23]. There were four hybrid loss functions are created. They are the classification loss (L_{cls}), the distangled cosine distance loss (L_{cos}), the reconstruct loss (L_{rec}) and the KL loss (L_{kl}). Each functions formula is given under Eq. (6), (7), (8) and (9).

$$L_{kl} = KL(Y_{pp}|N(0,1)) + KL(Y_{pv}|N(0,1)) + KL(Y_{qq}|N(0,1)) + KL(Y_{qv}|N(0,1)) \quad (6)$$

$$L_{rec} = \|B_{1'} - B_1\|_2 + \|B_{2'} - B_2\|_2 \quad (7)$$

$$L_{cos} = \frac{Y_{pv} \times Y_{qv}}{\|Y_{pv}\| \times \|Y_{qv}\|} \quad (8)$$

$$L_{cls} = -x \times \log(D(B_n)) \quad (9)$$

Where, x represents the one hot vector in the truth table.

E. Classification

The Convolution Neural Network (CNN) classifier identify the input images [24]. The visual representations are effectively examined by its multi-layered method, which also eliminates any necessary elements. Four layers make up the CNN classifier: input image, convolutional layer, pooling layer, fully connected layer and output. Through instruction, CNN is the instance that operates the quickest. The magnitudes of all input images' must be of identical dimensions. Fig. 4 depicts the CNN architecture.

7) *Convolutional layer*: After compiling a short sample of images, the convolution layer leverages all of the layers to examine the complexity of every image it gets. It has a strong relationship with the characteristics of the displayed photos. It is given in Eq. (10),

$$X_i = \sum Y_j * M_k + A_s \quad (10)$$

A_s - represents the bias term, $*$ represents the convolutional operator, k^{th} filter convolutes by local region of Y_j called receptive field, X_i is the output of every filter.

8) *Pooling layer*: This layer limits the type and severity that the downstream sampling layer can use. The Pooling layer decreases the quantity of constraints, the feature map's computation quality and scale, training duration and excessive fitting. It is proceeded in Eq. (11) and Eq. (12),

$$\tilde{x}_1 = \frac{x_1 - \tilde{m}}{\tilde{p}} + 1 \quad (11)$$

$$\tilde{x}_2 = \frac{x_2 - \tilde{m}}{\tilde{p}} + 1 \quad (12)$$

X is the output layer; p is the pooling factor.

9) *Fully connected layer*: A fully connected layer has been employed to categorize the images. All of the following convolution layers are placed after the FC layers. Arranging the graphical representation amongst both the input and the output is made simpler by the FC layer. The top layers of the network are fully connected layers. The layer that is completely connected receives its input from the pooling layer's output. The algorithm for the proposed DL-VAE-CNN is assumed below and the flow chart for the planned methodology is displayed in Fig. 5.

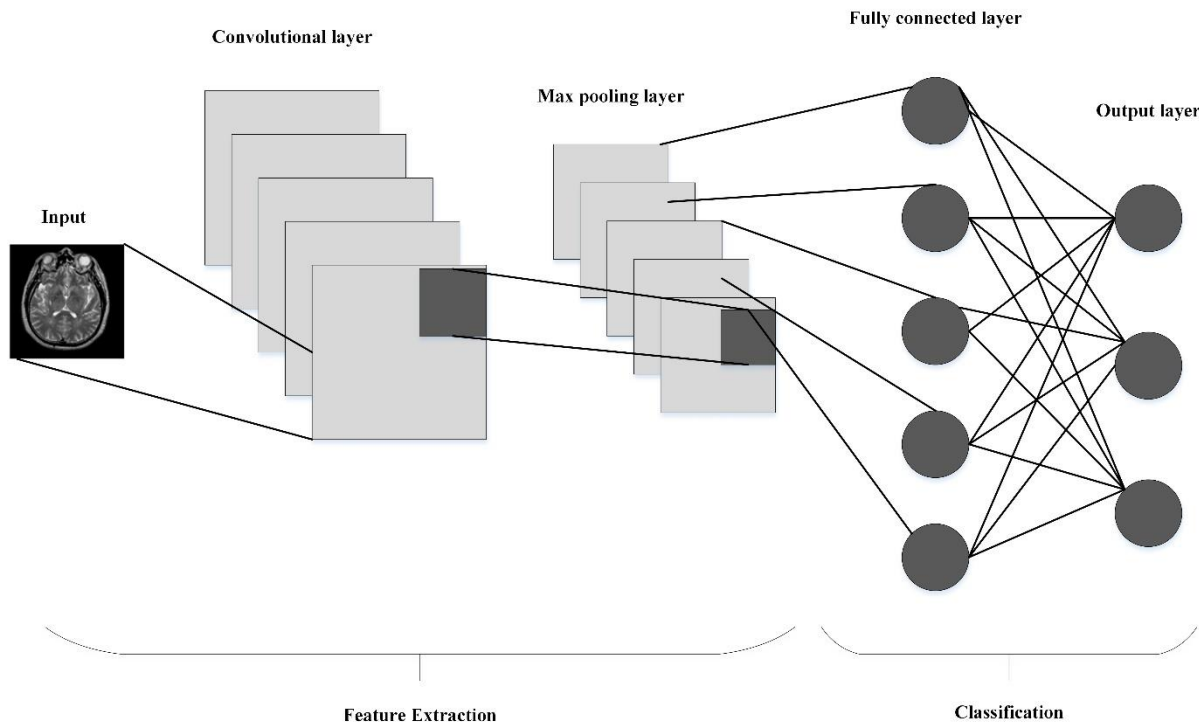


Fig. 4. Convolutional neural network architecture.

The framework's segmentation accuracy, efficiency, and resilience are greatly influenced by the choice and setup of parameters including learning rates, batch sizes, network designs, and hyper parameters. The capacity of the algorithm to handle various medical pictures, account for changes in image quality, and react to certain clinical settings may be dramatically impacted by fine-tuning these parameters. Additionally, the parameter selection may alter the amount of computer resources needed, reducing the algorithm's suitability for use in actual clinical settings. Therefore, further investigation and optimization of these parameters is necessary to guarantee the DL-VAE-CNN algorithm's optimal performance and clinical applicability for automated medical picture segmentation tasks.

Algorithm for DL-VAE-CNN

Input: Medical image of brain

Output: Classification of medical image

$Y = \{Y_1, Y_2, Y_3, \dots\}$ // LGG segmentation dataset

Image Pre-processing

Apply Histogram Equalization to enhance image contrast

Histogram Equalization equation is given by Eqn. (1)

Image Segmentation using K-means Clustering

Detect the cluster centres using the K-means algorithm

Calculate Euclidean distance E_d as in Eqn. (2)

Assign pixels to the nearest cluster centre using E_d

Update cluster centres using the recalculated coordinates in Eqn. (3)

Repeat until convergence

Feature Extraction using VAE

Implement the VAE architecture

Encoder network with Eqn. (4) and (5) for mean and variance

Sample latent variables using the parameterization trick

Decoder network reconstructs the input image

Calculate loss functions as in Eqn. (6), (7), (8), and (9)

Train the VAE using back propagation and optimization

Classification using CNN

Implement a CNN architecture for classification

Apply convolutional layers with Eqn. (11) and (12) for feature extraction

Apply pooling layers to down sample the feature maps

Flatten the feature maps and connect to fully connected layers

Implement softmax layer for multi-class classification

Training and Evaluation

Update VAE and CNN parameters iteratively through back propagation

Monitor loss functions and classification accuracy during training

Inference

Feed an unseen medical image through the trained DL-VAE-CNN framework

The framework will automatically segment and classify the image

Output

The DL-VAE-CNN algorithm outputs a classification label for the input medical image

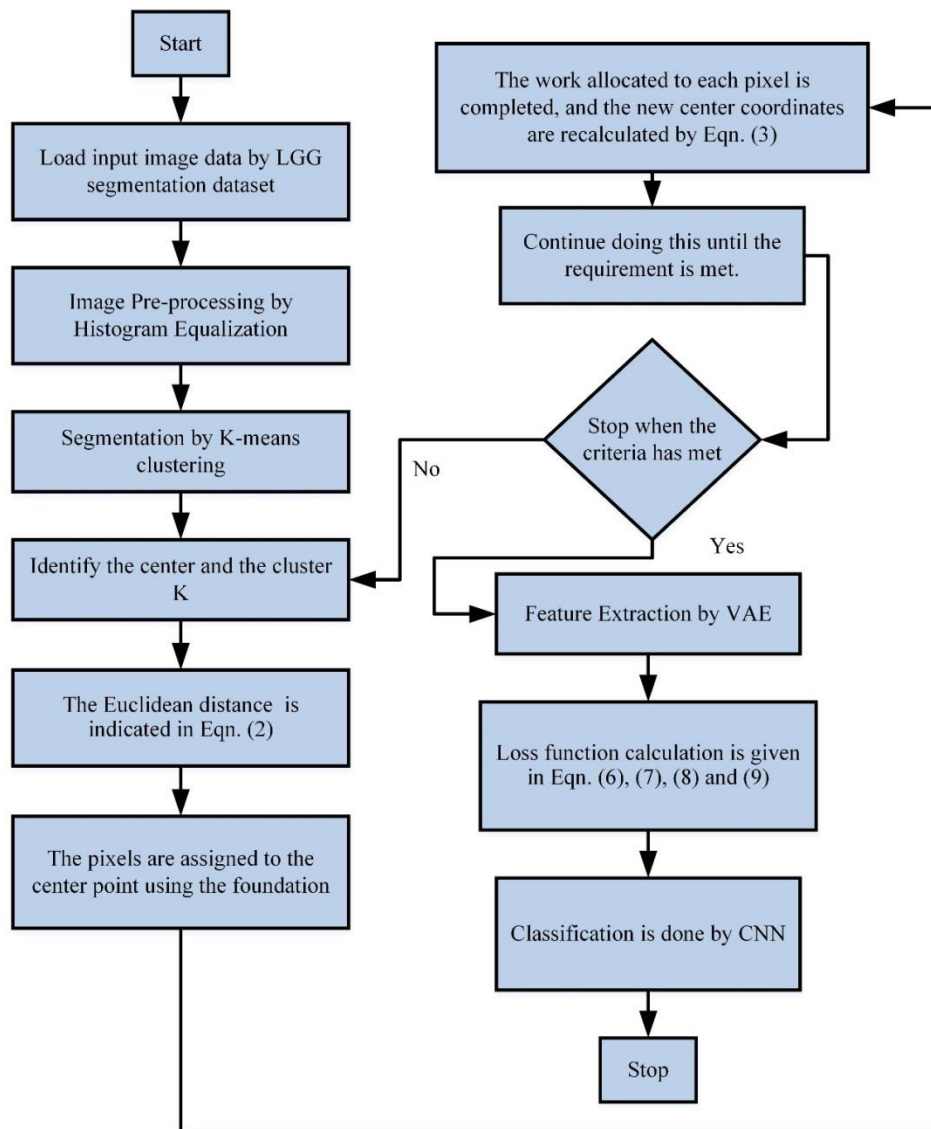


Fig. 5. Flow chart of proposed DL-VAE-CNN model.

V. RESULTS AND DISCUSSIONS

Deep learning, variational autoencoders and conditional neural networks were used to attain highly precise results for segmentation in the suggested automated medical image segmentation methodology. Comparison with existing methods revealed considerable gains in segmentation clarity as indicated by measures like precision, accuracy, fl score and recall. By efficiently employing the previously acquired hidden representations from variational autoencoders, the structure demonstrated adaptability to changes in the quality of images, especially changes in illumination, contrary and noise. By using condensed latent visualizations, the system also showed that it is capable of segmenting data quickly and effectively. The framework's generalization capabilities were especially outstanding because they allowed it to deal with a variety of anatomical characteristics and modalities for medical imaging with ease. These outcomes demonstrate that the suggested methodology for computerized medical image segmentation is efficient.

A. Performance Metrics Evaluation

1) *Accuracy*: The percentage of uniqueness among a calculation and its honest worth is known as accuracy. It is the ratio of precisely intended information to all observations. It is given in Eq. (13).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

2) *Precision*: The level of accuracy or closeness between multiple computations is referred to as precision. The relationship between precision and accuracy reveals how reproducible the measurement is. Its formula is given in Eq. (14).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (14)$$

3) *Recall*: The ratio of all suitable results that were precisely organized by the method used is known as recall. The correct affirmative to true positive and false negative

values proportion is used to characterize it. It is utilized in Eq. (15).

$$\text{Recall} = \frac{TP}{TP+FN} \quad (15)$$

4) *F1-Score*: Precision and recall are mutual to regulate the F1-score. When computing the F1 score, precision and recall levels are extremely important. Its formula is given in Eq. (16).

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

Table I and Fig. 6 shows the assessment and performance evaluation of Accuracy. When comparing the proposed DL-VAE-CNN methods accuracy parameter with the subsequent three existing methods, i) CNN procedure [25], ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater accuracy of about (99.91%).

TABLE I. COMPARISON TABLE OF PROPOSED METHOD'S ACCURACY WITH EXISTING METHOD'S ACCURACY

Method	Accuracy (%)
CNN	95.44
VGG16-CNN	93.74
DCNN	98.51
Proposed Method	99.91

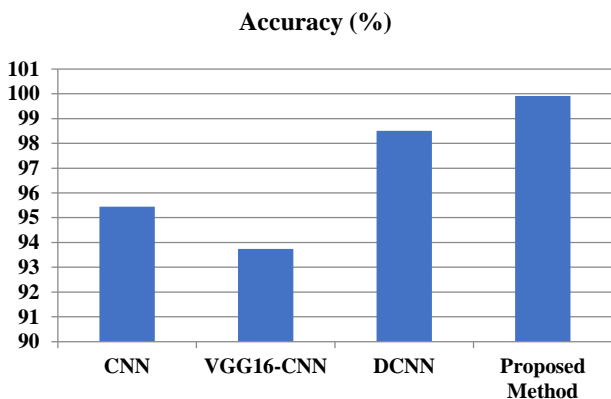


Fig. 6. Comparison graph of accuracy.

Table II and Fig. 7 shows the assessment and performance evaluation of Precision. When comparing the proposed DL-VAE-CNN methods precision parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater precision of about (99.90%).

TABLE II. COMPARISON TABLE OF PROPOSED METHOD'S PRECISION WITH EXISTING METHOD'S PRECISION

Method	Precision (%)
CNN	91
VGG16-CNN	92
DCNN	99.18
Proposed Method	99.90

Precision (%)

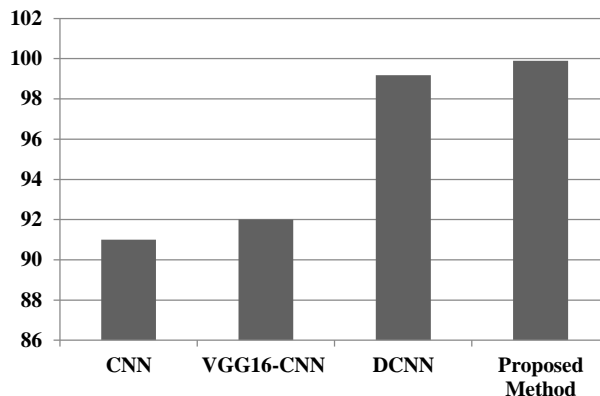


Fig. 7. Comparison graph of precision.

Table III and Fig. 8 shows the assessment and performance evaluation of Recall. When comparing the proposed DL-VAE-CNN methods recall parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater recall of about (98.99%).

TABLE III. COMPARISON TABLE OF PROPOSED METHOD'S RECALL WITH EXISTING METHOD'S RECALL

Method	Recall (%)
CNN	95
VGG16-CNN	92.1
DCNN	97.90
Proposed Method	98.99

Recall (%)

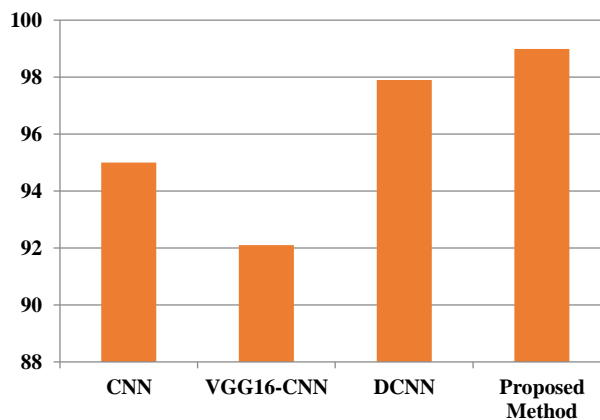


Fig. 8. Comparison graph of recall.

Table IV and Fig. 9 shows the assessment and performance evaluation of F1-Score. When comparing the proposed DL-VAE-CNN methods f1-score parameter with the subsequent three existing methods, i) CNN procedure [25] ii) VGG16-CNN [26], iii) DCNN [27], the proposed algorithm produces greater f1-score of about (99.99%).

TABLE IV. COMPARISON TABLE OF PROPOSED METHOD'S F1-SCORE WITH EXISTING METHOD'S F1-SCORE

Method	F1-Score (%)
CNN	93
VGG16-CNN	67.08
DCNN	98.53
Proposed Method	99.99

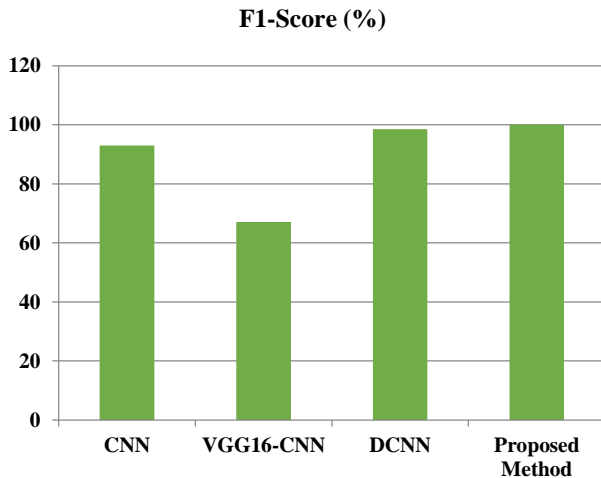


Fig. 9. Comparison graph of F1-score.

B. Discussions

An automated medical image segmentation framework using Deep Learning and Variational Auto Encoder with Convolutional Neural Network is employed in this paper. This section explains about the things that have explained above in a short way. Initially, the introduction, related works, problem statements are completed. Next, it steps on to the methodology part. Here, it discusses about the planned method that is used in this paper. Fig. 1 shows the architectural illustration of the planned DL-VAE-CNN technique. Fig. 2 shows the Histogram Equalization of the healthcare image. Next, the segmentation process takes place and it is done by K-means clustering. It is followed by that the feature extraction has done. Fig. 3 shows the architectural diagram of VAE and followed by that its components are explained. Next, is the classification process. It is done by CNN. Fig. 4 shows the architectural diagram of CNN and followed by that the various layers of CNN has been explained. Next, is the algorithm of DL-VAE-CNN. Fig. (5) shows the flow chart of DL-VAE-CNN. Next, is the results and discussion's part. Here, the various performance metrics like Precision, Accuracy, F1-score and Recall are associated with existing methods to show the planned method will produce better results. Table I gives the comparison table of accuracy and Fig. 6 shows the evaluation graph of accuracy. Table II gives the comparison table of precision and Fig. 7 shows the evaluation graph of precision. Table III gives the comparison table of recall and Fig. 8 shows the evaluation graph of recall. Table IV gives the comparison table of f1-score and Fig. 9 shows the evaluation graph of f1-score. Finally, the discussion part and followed by that conclusion and the future work of the paper is provided.

VI. CONCLUSION AND FUTURE WORK

In the field of medical imaging, establishing trustworthy image authentication to establish correlations between images is a difficult and critical task that is addressed in this study. The generation of tissue atlases, tumor development tracking, and image fusion are just a few therapeutic applications where this association is crucial. The paper offers a fresh method for addressing this issue, using CNNs and VAEs for healthcare data separation as deep learning approaches. Due to the requirement for precise image division at a low-level spatial data level, the fundamental machine vision task of image segmentation is extremely difficult. The suggested approach uses CNNs and VAEs in a conditioned environment to integrate latent representations that VAEs can produce using. When the trained model delineates pertinent regions in recently collected medical images during the inferential step, it successfully demonstrates its efficacy. The experimental findings drawn from several medical imaging databases highlight the increased accuracy in segmenting anatomical entities. This accuracy highlights the substantial contribution of this method to the field of medical imaging and expands the possibilities of automated diagnosis and therapy. The complex pixel-level classification issue, which has presented difficulties in earlier studies, is effectively addressed by the suggested Deep Learning and DL-VAE-CNN approach. This research provides a strong framework that has the potential to revolutionize image identification and segmentation in the context of medical imaging by using the capabilities of deep learning, latent space representation, and conditional neural networks. Deep learning frameworks sometimes require a lot of processing power, both during inference and training. Deploying it in clinical settings with limited resources may thus be difficult.

REFERENCES

- [1] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, "Brain Tumor Classification Using a Combination of Variational Autoencoders and Generative Adversarial Networks," *Biomedicine*, vol. 10, no. 2, p. 223, Jan. 2022, doi: 10.3390/biomedicine10020223.
- [2] W. T. Le, F. Maleki, F. P. Romero, R. Forghani, and S. Kadoury, "Overview of Machine Learning: Part 2," *Neuroimaging Clin. N. Am.*, vol. 30, no. 4, pp. 417–431, Nov. 2020, doi: 10.1016/j.nic.2020.06.003.
- [3] N. Sharma et al., "Automated medical image segmentation techniques," *J. Med. Phys.*, vol. 35, no. 1, p. 3, 2010, doi: 10.4103/0971-6203.58777.
- [4] V. Sandfort, K. Yan, P. M. Graffy, P. J. Pickhardt, and R. M. Summers, "Use of Variational Autoencoders with Unsupervised Learning to Detect Incorrect Organ Segmentations at CT," *Radiol. Artif. Intell.*, vol. 3, no. 4, p. e200218, Jul. 2021, doi: 10.1148/ryai.2021200218.
- [5] H. Uzunova, S. Schultz, H. Handels, and J. Ehrhardt, "Unsupervised pathology detection in medical images using conditional variational autoencoders," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 3, pp. 451–461, Mar. 2019, doi: 10.1007/s11548-018-1898-0.
- [6] R. Wei and A. Mahmood, "Recent Advances in Variational Autoencoders With Representation Learning for Biomedical Informatics: A Survey," *IEEE Access*, vol. 9, pp. 4939–4956, 2021, doi: 10.1109/ACCESS.2020.3048309.
- [7] P. Naga Srinivasu, T. B. Krishna, S. Ahmed, N. Almusallem, F. Khaled Alarfaj, and N. Allheeb, "Variational Autoencoders-Based Self-Learning Model for Tumor Identification and Impact Analysis from 2-D MRI Images," *J. Healthc. Eng.*, vol. 2023, pp. 1–17, Jan. 2023, doi: 10.1155/2023/1566123.
- [8] P. Celard, E. L. Iglesias, J. M. Sorribes-Fdez, R. Romero, A. S. Vieira, and L. Borrajo, "A survey on deep learning applied to medical images:

- from simple artificial neural networks to generative models,” *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2291–2323, Jan. 2023, doi: 10.1007/s00521-022-07953-4.
- [9] I. Rizwan I Haque and J. Neubert, “Deep learning approaches to biomedical image segmentation,” *Inform. Med. Unlocked*, vol. 18, p. 100297, 2020, doi: 10.1016/j.imu.2020.100297.
- [10] X. Chen et al., “Recent advances and clinical applications of deep learning in medical image analysis,” *Med. Image Anal.*, vol. 79, p. 102444, Jul. 2022, doi: 10.1016/j.media.2022.102444.
- [11] Z.-S. Liu, W.-C. Siu, and Y.-L. Chan, “Photo-Realistic Image Super-Resolution via Variational Autoencoders,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 4, pp. 1351–1365, Apr. 2021, doi: 10.1109/TCSVT.2020.3003832.
- [12] H. Uzunova, J. Ehrhardt, T. Kepp, and H. Handels, “Interpretable explanations of black box classifiers applied on medical images by meaningful perturbations using variational autoencoders,” in *Medical Imaging 2019: Image Processing*, E. D. Angelini and B. A. Landman, Eds., San Diego, United States: SPIE, Mar. 2019, p. 36. doi: 10.1117/12.2511964.
- [13] W. Yu, M. Zhang, and Y. Shen, “Spatial Revising Variational Autoencoder-Based Feature Extraction Method for Hyperspectral Images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1410–1423, Feb. 2021, doi: 10.1109/TGRS.2020.2997835.
- [14] G. Peyré and M. Cuturi, *Computational Optimal Transport*. now Publishers Inc, 2019. doi: 10.1561/9781680835519.
- [15] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image Segmentation Using Deep Learning: A Survey.” *arXiv*, Nov. 14, 2020. Accessed: Jul. 06, 2023. [Online]. Available: <http://arxiv.org/abs/2001.05566>
- [16] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, “Medical image segmentation using deep learning: A survey,” *IET Image Process.*, vol. 16, no. 5, pp. 1243–1267, Apr. 2022, doi: 10.1049/ipr2.12419.
- [17] V. Edupuganti, M. Mardani, S. Vasanawala, and J. Pauly, “Uncertainty Quantification in Deep MRI Reconstruction,” *IEEE Trans. Med. Imaging*, vol. 40, no. 1, pp. 239–250, Jan. 2021, doi: 10.1109/TMI.2020.3025065.
- [18] T. Le-Tien, T.-N. To, and G. Vo, “Graph-based signal processing to convolutional neural networks for medical image segmentation,” *SEATUC J. Sci. Eng.*, vol. 3, no. 1, pp. 9–15, 2022.
- [19] R. J. S. Raj, S. J. Shobana, I. V. Pustokhina, D. A. Pustokhin, D. Gupta, and K. Shankar, “Optimal Feature Selection-Based Medical Image Classification Using Deep Learning Model in Internet of Medical Things,” *IEEE Access*, vol. 8, pp. 58006–58017, 2020, doi: 10.1109/ACCESS.2020.2981337.
- [20] Dr. S. Manoharan, “Performance Analysis of Clustering Based Image Segmentation Techniques,” *J. Innov. Image Process.*, vol. 2, no. 1, pp. 14–24, Mar. 2020, doi: 10.36548/jiip.2020.1.002.
- [21] M. Imani and H. Ghassemian, “An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges,” *Inf. Fusion*, vol. 59, pp. 59–83, Jul. 2020, doi: 10.1016/j.inffus.2020.01.007.
- [22] K.-L. Lim, X. Jiang, and C. Yi, “Deep Clustering With Variational Autoencoder,” *IEEE Signal Process. Lett.*, vol. 27, pp. 231–235, 2020, doi: 10.1109/LSP.2020.2965328.
- [23] Q. Zuo, Y. Zhu, L. Lu, Z. Yang, Y. Li, and N. Zhang, “Fusing Structural and Functional Connectivities using Disentangled VAE for Detecting MCI.” *arXiv*, Jun. 16, 2023. Accessed: Jul. 07, 2023. [Online]. Available: <http://arxiv.org/abs/2306.09629>
- [24] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges,” *J. Digit. Imaging*, vol. 32, no. 4, pp. 582–596, Aug. 2019, doi: 10.1007/s10278-019-00227-x.
- [25] M. A. Mahjoubi, S. Hamida, O. E. Gannour, B. Cherradi, A. E. Abbassi, and A. Raihani, “Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140346.
- [26] A. Alshammari, “Construction of VGG16 Convolution Neural Network (VGG16_CNN) Classifier with NestNet-Based Segmentation Paradigm for Brain Metastasis Classification,” *Sensors*, vol. 22, no. 20, p. 8076, Oct. 2022, doi: 10.3390/s22208076.
- [27] L. Alzubaidi et al., “Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data,” *Cancers*, vol. 13, no. 7, p. 1590, Mar. 2021, doi: 10.3390/cancers13071590.

Estimating Probability Values Based on Naïve Bayes for Fuzzy Random Regression Model

Hamijah Mohd Rahman¹, Nureize Arbai², Chuah Chai Wen³, Pei-Chun Lin⁴

Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia, Johor, Malaysia^{1,2,3}
Dept. of Information Engineering and Computer Science, Feng Chia University, Taichung, Taiwan⁴

Abstract—In the process of treating uncertainties of fuzziness and randomness in real regression application, fuzzy random regression was introduced to address the limitation of classical regression which can only fit precise data. However, there is no systematic procedure to identify randomness by means of probability theories. Besides, the existing model mostly concerned in fuzzy equation without considering the discussion on probability equation though random plays a pivotal role in fuzzy random regression model. Hence, this paper proposed a systematic procedure of Naïve Bayes to estimate the probabilities value to overcome randomness. From the result, it shows that the accuracy of Naïve Bayes model can be improved by considering the probability estimation.

Keywords—Naïve Bayes; fuzziness; randomness; probability estimation

I. INTRODUCTION

Fuzziness and randomness are two uncertainties involved in practice of real observation where the statistical data are collected from various measurements. Fuzziness comes from incomplete information while randomness can be related to stochastic variability of all possible outcomes of a situation [1]. In mathematical viewpoint, both uncertainties are merged to formulate a fuzzy random variable by means of assigning probability and fuzzy set theories since possible random outcome have to be described by terms of fuzzy set.

Fuzzy random variable had been studied by many researchers over past few decades. The first introduction of fuzzy random variable concept had been given by Kwarkernaak [2][3]. Since then, different researchers studied fuzzy random variable according to different requirements like by Puri and Ralescu [4] and Liu and Liu [5]. Considering the ability of fuzzy random theories in handling simultaneously fuzzy random uncertainties, this approach can be found in various applications such as in regression analysis. Nather [6] presented fuzzy random variable to deal with regression analysis when the statistical has linguistic data.

In the situation where randomness and fuzziness associated in the regression problems, fuzzy random regression was introduced as a solution for real life regression analysis where the data is not only characterized by imprecision and vagueness but there also exist the formalism of random variables. Fuzzy random regression based on fuzzy random variables with confidence interval was proposed in the framework of real regression analysis where there exist uncertainties [7]. The implementation of this technique as an integral component of regression was successfully in

achieving the objective to estimate weight in the production of oil palm [8]. Considering the statistical used content fuzzy random information, fuzzy random regression was proposed to estimate coefficient in the model setting [9]. In another application, fuzzy random was introduced to build an improved fuzzy random regression for data preparation by using time series data [10][11].

Various applications presented fuzzy random concept due to its capability in handling factors of fuzziness and randomness. However, the existing studies mostly focused on fuzzy equation regarded as the concept of possibility, without considering the probability equation. Moreover, the models are not adequately discussed on how to estimate probability to reduce randomness [6][7][11]. To date, probability theory is used to model randomness which recorded from dispersion of the measured value [9]. Hence, according to abovementioned reason, this study is to present a systematic procedure to control randomness. This study is concentrated on developing a procedure of probability estimation for the fuzzy random data. This systematic procedure is important to guide the identification of probability estimation in defining the fuzzy random data for developing fuzzy random regression model.

The remainder of this paper is arranged as follow. Some preliminaries of uncertainty fuzzy random is covered in Section II. Next section discusses the procedure of proposed method that is Naïve Bayes to estimate random value for fuzzy random regression model. An empirical study is provided in Section IV to illustrate the proposed method. Finally, Section V discusses the conclusion.

II. THEORETICAL BACKGROUND

A. Fuzzy Random

The fuzzy random variable was introduced to present the real situation of uncertainty which comes from vagueness, imprecision, randomness etc. [1]-[5]. The concept of fuzzy random variable has been applied in several papers which combine both fuzzy random uncertainties [18]-[22].

Definition 1. Let Y be the fuzzy variable with possibility distribution μ_Y , the possibility, necessity, and credibility of event $\{Y \leq r\}$ are given in equation as follows.

$$Pos\{Y \leq r\} = \sup \mu_Y(t), t \leq r \quad (1)$$

$$Nec\{Y \leq r\} = 1 - \sup(t), t \geq r \quad (2)$$

$$Cr\{Y \leq r\} = \frac{1}{2}(1 + \sup_{t \leq r} \mu_Y(t) - \sup_{t \leq r} \mu_Y(t)) \quad (3)$$

where,

Cr = Credibility measure

Pos = possibility

Nec = necessity measure

Credibility measure is an average of the possibility and the necessity measures, $Cr\{.\} = (Pos\{.\} + Nec\{.\})/2$. Credibility measure is presented to expand a certain measure of possibility and necessity, which is a sound of aggregate of two cases.

Definition 2: Let Y be a fuzzy variable. Under the assumption that the two integrals are finite, the expected value of Y is defined as follows in (7):

$$E[Y] = \int_0^x Cr\{Y \geq r\}dr - \int_{-\infty}^0 Cr\{Y \leq r\}dr \quad (4)$$

Following from Equation (4), the expected value of Y is defined as

$$E[Y] = \frac{a^l + 2c + a^r}{4} \quad (5)$$

where $Y = (c, a^l, a^r)_T$ is a triangular fuzzy number and c is a center value.

The expected value of the fuzzy variable $X(\omega)$ is denoted by $E[X(\omega)]$ for any fuzzy random variable X on Ω [5]. Thus, the expected value of the fuzzy random variable X is defined as the mathematical expectation of the random variable $E[X(\omega)]$.

Definition 3: Let X be a fuzzy random variable defined on a probability space (Ω, Σ, Pr) with expected value e . The expected value of X is defined in Equation (6) as follows.

$$E[\xi] = \int \Omega [\int_0^\infty Cr\{\xi(\omega) > r\}dr - \int_{-\infty}^0 Cr\{\xi(\omega) \leq r\}dr]Pr(d\omega) \quad (6)$$

The variance of X [7] is defined as Equation (7), respectively.

$$Var[X] = E[(X - e)^2] \quad (7)$$

where $e = E[X]$.

B. Naïve Bayes

The Naïve Bayes is a simplification of Bayes Theorem which is used as a classification algorithm with an assumption of independence among predictors [8]. It is known as 'Naïve' because it assumes that the presence of input features is independent of each other. As the feature of the data points is unrelated to any other, therefore, changing of one input feature may not affect others [9].

The general equation for Bayes [10][11] is given as follows:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (8)$$

where,

$P(A)$ = the probability of A occurring

$P(B)$ = the probability of B occurring

$P(A|B)$ = the probability of A given B

$P(B|A)$ = the probability of B given A

$P(A \cap B)$ = the probability of both A and B occurring

Given A as Hypothesis and B as evidence, the Bayes rules derive the probability of a hypothesis given the evidence. The rule stated the relationship incorporating $P(A)$ distribution in order to generate $P(A|B)$. $P(A)$ is the probability of an event before getting the evidence. The probability of the event based on the current knowledge before an experiment is performed. $P(A|B)$ is called the posterior probability which is calculated by updating the prior probability after taking into consideration new information. Meaning that, the posterior probability is the probability of event A occurring given that even B has occurred.

In this study, the Naïve Bayes theorem was proposed to be used in estimating probability values in constructing confidence intervals for fuzzy random regression model. This probability value is representative of randomness in the fuzzy random regression model. However, most studies [12] – [17] do not clearly describe how to obtain these random values to develop fuzzy random regression models. But these random values are very necessary to manage data that have random and fuzzy uncertainties to create forecasting models. The following section describes the proposed procedure for estimating probabilities for random values using the Naïve Bayes method that will be used to develop a fuzzy random regression prediction model.

III. ESTIMATING PROBABILITIES FOR FUZZY RANDOM REGRESSION PROCEDURE

This section describes the standard procedure to estimate probability value in developing confidence interval for Fuzzy Random Regression model. The procedure uses Naïve Bayes to characterize the random uncertainty.

The procedure for implementing the proposed method can be written in the following to determine the probabilities by using Naïve Bayes.

- Step 1 : Suppose x_n are the fuzzy data. Transform crisp data x_n into fuzzy random data (FRD), c_n, a_n^l, a_n^r
- Step 2 : Estimate probabilities for each FRD by using Naïve Bayes approach based on Equation (8). The fuzzy random data with probabilities can be arranged in the format as in Table I.

TABLE I. DATA FORMAT FOR FUZZY RANDOM DATA

Sample	Input	FRD	Pr
1	X_1	c_1, a_1^l, a_1^r	Pr_1
2	X_2	c_2, a_2^l, a_2^r	Pr_2
...
n	X_n	c_n, a_n^l, a_n^r	Pr_n

Step 3 : Calculate the expected value, $E(x)$ using the center of triangular fuzzy variable with probability which $Pr + Pr = 1$. The formulation to calculate the expected value is shown in Equation (9) as follows:

$$E(x) = Pr_{v_1}.E[x(V_1)] + Pr_{v_2}.E[x(V_2)] \quad (9)$$

Step 4 : Calculate variance, $Var(x)$. Define the variance of x by using Equation (7). The $E[(x(v_1) - E[x])^2]$ should be calculated to obtain the $Var(x)$. The calculation to obtain the variance is based on equation (10) respectively.

$$Var(x) = E[(x(v_1) - E[x])^2].Pr_{v_1} + E[(x(v_2) - E[x])^2].Pr_{v_2} \quad (10)$$

Step 5 : Determine the confidence interval, CI of FRD using Equation (11) as follows.

$$CI = [(E(x) - Var(x)), (E(x) + Var(x))] \quad (11)$$

Step 6 : Estimate coefficient based on confidence interval in Step 5. The coefficient can be obtained using the following linear programming.

$$\min J(\tilde{A}) = \sum_{k=1}^K (\tilde{A}_k^r - \tilde{A}_k^l) \quad \text{subject to} \quad \tilde{A}_k^r \geq \tilde{A}_k^l \quad (12)$$

$$\tilde{Y}_i = \sum_{k=1}^K \tilde{A}_k 1[ex_{ik}, \sigma x_{i1}] \supseteq_h I[eY_i, \sigma Y_i] \quad i = 1, \dots, n; 1, \dots, K,$$

The fuzzy random regression [7] prediction model was introduced with the advantage of handling data that had dual uncertainties namely fuzziness and randomness. Although this model is good for handling uncertainties, there are constraints for the industry to apply this model in the real world if there is no complete method specially to transform normal data into an acceptable form of data by this model. Then the standard method that has been tested has been introduced [18] – [20]. However, some of them focus on methods of managing fuzzy data only. Thus, this study specializes in the development of standard procedures for determining random value for the development of fuzzy random models.

IV. NUMERICAL EXPERIMENT

In this section, a numerical experiment has demonstrated to visualize how the probabilities are estimated using proposed procedure in order to handle randomness. The fuzzy random input and output data are taken from [7] in Table II and Table III, respectively.

Table II shows the fuzzy random input data with two attributes (X_1, X_2). Each attribute has four samples which

divided into center, left and right (c, a^l, a^r). Table III shows the fuzzy random output data with four samples of attribute Y and divided into (c, a^l, a^r).

TABLE II. FUZZY RANDOM INPUT DATA

Sample	X	FRD1			FRD2		
		c	a ^l	a ^r	c	a ^l	a ^r
1	X1 ₁	3	2	4	4	3	5
2	X1 ₂	6	4	8	8	6	10
3	X1 ₃	12	10	14	14	12	16
4	X1 ₄	14	12	16	16	14	18
Sample	X	FRD1			FRD2		
		c	a ^l	a ^r	c	a ^l	a ^r
1	X2 ₁	2	1	3	4	3	5
2	X2 ₂	3	2	4	4	3	5
3	X2 ₃	12	10	16	14	12	16
4	X2 ₄	18	16	20	21	20	22

TABLE III. FUZZY RANDOM OUTPUT DATA

Sample	X	FRD1			FRD2		
		c	a ^l	a ^r	c	a ^l	a ^r
1	Y ₁	14	10	16	18	16	20
2	Y ₂	17	16	18	20	18	22
3	Y ₃	22	20	24	26	24	28
4	Y ₄	32	30	34	36	32	40

By using the values in fuzzy random data, the probabilities for each value can be determined. Using calculation in Equation (8), the probabilities for each fuzzy random data are estimated.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

Assuming each feature variable is independent of the rest, calculate the probability of each separate feature given of each class. First step is finding the prior probability of each class in X_1 . Given sample X_1 has four variables ($X_{1_1}, X_{1_2}, X_{1_3}, X_{1_4}$). Each variable has one event occur. In mathematical, the probability can be represented as $P(X_{1_1}) = \frac{1}{4} = 0.25$ which mean the occurring event happening is likely one time by considering the total of potential outcome. Thus, the prior probability of each class in X_1 is as follows.

$$\begin{aligned} P(X_{1_1}) &= \frac{1}{4} = 0.25 \\ P(X_{1_2}) &= \frac{1}{4} = 0.25 \\ P(X_{1_3}) &= \frac{1}{4} = 0.25 \\ P(X_{1_4}) &= \frac{1}{4} = 0.25 \end{aligned} \quad (13)$$

Following the calculation in Equation (13), find probabilities for input X_2 and output Y . The result for each probability is tabulated in Table IV and Table V, respectively.

Table IV and Table V show the probabilities for fuzzy random input and output data. The probabilities for FRD1, Pr_1 are determined using calculation in (13), with value 0.25. Note that, probability is counted by $Pr_1 + Pr_2 = 1$. Therefore, the probability value for FRD2, Pr_2 is 0.75. These probabilities are used to calculate the expected value $E(x)$ using the center of triangular fuzzy variable as in Equation (9). Based on these values of $E(x)$, variance $Var(x)$ can be calculated using Equation (10). The results for $E(x)$ and $Var(x)$ are tabulated in Table VI respectively.

TABLE IV. FUZZY RANDOM DATA WITH PROBABILITIES FOR INPUT DATA

Sample	X1	FRD1				FRD 2			
		c	a ^l	a ^r	Pr ₁	c	a ^l	a ^r	Pr ₂
1	X1 ₁	3	2	4	0.25	4	3	5	0.75
2	X1 ₂	6	4	8	0.25	8	6	10	0.75
3	X1 ₃	12	10	14	0.25	14	12	16	0.75
4	X1 ₄	14	12	16	0.25	16	14	18	0.75
Sample	X2	FRD1				FRD 2			
		c	a ^l	a ^r	Pr ₁	c	a ^l	a ^r	Pr ₂
1	X2 ₁	2	1	3	0.25	4	3	5	0.75
2	X2 ₂	3	2	4	0.25	4	3	5	0.75
3	X2 ₃	12	10	16	0.25	14	12	16	0.75
4	X2 ₄	18	16	20	0.25	21	20	22	0.75

TABLE V. FUZZY RANDOM DATA WITH PROBABILITIES FOR OUTPUT DATA

Sample	Y	FRD1				FRD 2			
		c	a ^l	a ^r	Pr ₁	c	a ^l	a ^r	Pr ₂
1	Y ₁	14	10	16	0.25	18	16	20	0.75
2	Y ₂	17	16	18	0.25	20	18	22	0.75
3	Y ₃	22	20	24	0.25	26	24	28	0.75
4	Y ₄	32	30	34	0.25	36	32	40	0.75

Table VI shows the value of expectation and variance for the input output fuzzy random data. The expectation and variance values are used to find confidence interval by using Equation (11). The results are tabulated in Table VII.

Table VII shows the confidence interval result for fuzzy random input and output data. In this study, the confidence interval was considered as one-sigma confidence ($1 \times \sigma$) interval of each fuzzy random variable. The combination of expectation and variance of fuzzy random variable was induced to define the confidence-interval-based-inclusion [7]. Based on this confidence interval, a fuzzy random regression model can be formulated using mathematical linear programming as in Equation (12) in order to define coefficient.

TABLE VI. EXPECTATION AND VARIANCE OF THE DATA

i	Ex ₁ , V _{x1}		Ex ₂ , V _{x2}		E _y , V _y	
1	3.75	0.5729	3.5	1.2031	16.2	10.6688
2	7.5	2.2917	3.75	0.5729	17.6	1.8113
3	13.5	2.2917	13.63	3.467	24.8	4.8125
4	15.5	2.2917	20.35	3.7138	34.4	4.8125

TABLE VII. CONFIDENCE INTERVAL FOR FUZZY RANDOM INPUT OUTPUT DATA

i	X1	X2	Y
1	[3.177, 4.323]	[2.297, 4.703]	[5.531, 26.869]
2	[5.208, 9.792]	[3.177, 4.323]	[15.789, 19.411]
3	[11.208, 15.792]	[10.158, 17.092]	[19.988, 29.613]
4	[13.208, 17.792]	[16.536, 23.964]	[29.588, 39.213]

$$\min J(\tilde{A}) = \sum_{k=1}^K (\tilde{A}_k^r - \tilde{A}_k^l)$$

subject to

$$\tilde{A}_k^r \geq \tilde{A}_k^l$$

$$\tilde{Y}_i = \sum_{k=1}^K \tilde{A}_k 1[ex_{ik}, \sigma x_{i1}] \supseteq_h I[ey_i, \sigma Y_i]$$

$$i = 1, \dots, n; 1, \dots, K,$$

$$\min = (a_1^r - a_1^l) + (a_2^r - a_2^l);$$

$$a_1^l \leq a_1^r;$$

$$3*a_1^l + 2*a_2^l \leq 3.177075;$$

$$6*a_1^l + 4*a_2^l \leq 5.208325;$$

$$12*a_1^l + 10*a_2^l \leq 11.208333;$$

$$14*a_1^l + 12*a_2^l \leq 13.208333;$$

$$3*a_1^r + 5*a_2^r \geq 4.322925;$$

$$6*a_1^r + 10*a_2^r \geq 9.791675;$$

$$12*a_1^r + 16*a_2^r \geq 15.791667;$$

$$14*a_1^r + 18*a_2^r \geq 17.791667;$$

$$a_1^l \geq 0; a_1^r \geq 0;$$

$$a_2^l \geq 0; a_2^r \geq 0; \tag{14}$$

The linear programming of the fuzzy random regression was applied to the dataset as shown in Equation (14). This linear programming is performed to generate the coefficient value as tabulated in Table VIII respectively.

TABLE VIII. COEFFICIENT OF THE FUZZY RANDOM INPUT OUTPUT DATA

Item	Coefficient		Width
	A1	A2	
Y	0.00	0.943	0.79
X1	0.00	1.044	0.113
X2	0.00	1.052	0.073

The coefficient result for the fuzzy random input output data as tabulated in Table VIII shows the values estimated from fuzzy random regression. The attribute which has larger coefficient value is more significant to the total evaluation. In this result, it shows that the evaluation of attributes A1 and A2 indicate the A2 is significant to the total evaluation due to its higher coefficient. The model had a wider coefficient width because of the consideration of the confidence interval in its evaluation. The width in this evaluation plays an important role, as it reflects natural human judgment.

A greater breadth denotes the evaluation's ability to capture more data while using fuzzy judgments. Mean Square Error (MSE) can be defined using the estimated coefficient obtained from the fuzzy random regression and the model in Equation (14). In Table IX, the mean squared error (MSE) is calculated to compare the outcomes of the existing approach and the suggested method.

TABLE IX. MSE RESULT

Watada [7]	Naïve Bayes
196.6845	193.8861

Table IX shows the MSE result using Naïve Bayes approach as compared with current method by Watada *et al.*, [7]. In comparison study, the testing data derived from proposed method have a close majority of the expectation and variance result when compare to [7]. As the majority of the expectation and variance have been captured, therefore, both confidence intervals from testing and current model are quite similar. The evaluation of MSE was considered using current model and testing. From the result shown in Table IX, MSE of the proposed model is smaller than the other. This MSE implies that the prediction error can be reduced significantly.

The outcomes of the experiment demonstrate that the suggested approach is highly accurate at estimating the expectation, variance, and confidence interval of the data. Additionally, it is more accurate than the present technique and has a lower MSE, proving its superiority. These findings imply that the suggested approach can estimate probability and circumvent data unpredictability.

V. CONCLUSION

In this paper, a procedure based on Naïve Bayes is proposed to treat data which contain uncertainty known as fuzzy random data. The uncertainty data of randomness was handled by implementing the Naïve Bayes method to estimate probability. As to demonstrate the potential application of proposed method for accessing estimation, an experimental study using fuzzy random data is illustrated and the results are compared with the result of current method. The result shows that the proposed method has majority close of the

expectation, variance and confidence interval. Further, it also has better MSE result than the current method. The result demonstrated that the proposed model is capable to estimate probability and overcome randomness of the data.

ACKNOWLEDGMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (Vot Q507) and Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/ICT02/UTHM/02/7). This research work is also supported by the Ministry of Education, R.O.C., under the grants of TEEP@AsiaPlus. The work of this paper is also supported by the Ministry of Science and Technology under Grant No. MOST 109-2221-E-035-063-MY2.

REFERENCES

- [1] Shapiro, A. F. (2013). Implementing Fuzzy Random Variables—Some Preliminary Observations. ARCH 2013.1 Proceedings, 1-15
- [2] Kwakernaak, H. (1978). Fuzzy random variables—I. Definitions and theorems. Information sciences, 15(1), 1-29.
- [3] Kwakernaak, H. (1979). Fuzzy random variables—II. Algorithms and examples for the discrete case. Information Sciences, 17(3), 253-278.
- [4] Puri, M. L., Ralescu, D. A., & Zadeh, L. (1993). Fuzzy random variables. In Readings in fuzzy sets for intelligent systems (pp. 265-271). Morgan Kaufmann.
- [5] Liu, Y. K., & Liu, B. (2003). Fuzzy random variables: A scalar expected value operator. Fuzzy Optimization and decision making, 2(2), 143-160.
- [6] Näther, W. (2006). Regression with fuzzy random data. Computational Statistics & Data Analysis, 51(1), 235-252.
- [7] Watada, J., Wang, S., & Pedrycz, W. (2009). Building confidence-interval-based fuzzy random regression models. IEEE Transactions on Fuzzy Systems, 17(6), 1273-1283.
- [8] Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics; Elsevier Science Publisher: Amsterdam, The Netherlands, 403-412.
- [9] Rouder, J. N., & Morey, R. D. (2018). Teaching Bayes' theorem: Strength of evidence as predictive accuracy. The American Statistician.
- [10] Davies, P. (1988). Kendall's Advanced Theory of Statistics. Volume 1. Distribution Theory. Nureize, A., Watada, J., & Wang, S. (2014). Fuzzy random regression based multi-attribute evaluation and its application to oil palm fruit grading. Annals of Operations Research, 219(1), 299-315.
- [11] Puga, J. L., Krzywinski, M., & Altman, N. (2015). Bayes' theorem: Incorporate new evidence to update prior information. Nature Methods, 12(4), 277-279.
- [12] Arbaiy, N., Watada, J., & Lin, P. C. (2016). Fuzzy random regression-based modeling in uncertain environment. In Sustaining power resources through energy optimization and engineering (pp. 127-146). IGI Global.
- [13] Efendi, R., Arbaiy, N., & Deris, M. M. (2018). A new procedure in stock market forecasting based on fuzzy random auto-regression time series model. Information Sciences, 441, 113-132.
- [14] Efendi, R., Samsudin, N. A., Arbaiy, N., & Deris, M. M. (2017, August). Maximum-minimum temperature prediction using fuzzy random auto-regression time series model. In 2017 5th International Symposium on Computational and Business Intelligence (ISCBI) (pp. 57-60). IEEE.
- [15] Hesamian, G., & Akbari, M. G. (2021). A robust multiple regression model based on fuzzy random variables. Journal of Computational and Applied Mathematics, 388, 113270.
- [16] Wang, T., Shi, P., & Wang, G. (2020). Solving fuzzy regression equation and its approximation for random fuzzy variable and their application. Soft Computing, 24(2), 919-933.
- [17] Shao, L., Tsai, Y. H., Watada, J., & Wang, S. (2012). Building Fuzzy Random Autoregression Model and Its Application. In Intelligent Decision Technologies (pp. 155-164). Springer, Berlin, Heidelberg.

- [18] Lah, M.S.C., Arbaiy, N. A simulation study of first-order autoregressive to evaluate the performance of measurement error based symmetry triangular fuzzy number, (2020) Indonesian Journal of Electrical Engineering and Computer Science, 18 (3), pp. 1559-1567.
- [19] Lah, M. S. C., Arbaiy, N., & Lin, P. C. (2020). Stock Index Modelling Using Arima With Standard Deviation Based Triangular Fuzzy Numbers. *Journal of Critical Reviews*, 7(8), 1264-1268.
- [20] Rahman, H. M., Arbaiy, N., Efendi, R., & Wen, C. C. (2019). Forecasting ASEAN countries exchange rates using auto regression model based on triangular fuzzy number. *J Electric Eng Comput Sci* 14 (3): 1525–1532.
- [21] González-De La Fuente, L., Nieto-Reyes, A., & Terán, P. (2022). Statistical depth for fuzzy sets. *Fuzzy Sets and Systems*, 443, 58-86.
- [22] Singh, V. P., Sharma, K., & Chakraborty, D. (2023). Solving capacitated vehicle routing problem with demands as fuzzy random variable. *Soft Computing*, 1-21.

A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data

Hairani Hairani, Dadang Priyanto

Department Computer Science, Bumigora University, Mataram, Indonesia

Abstract—The performance of machine learning methods in disease classification is affected by the quality of the dataset, one of which is unbalanced data. One example of health data that has unbalanced data is diabetes disease data. If unbalanced data is not addressed, it can affect the performance of the classification method. Therefore, this research proposed the SMOTE-ENN approach to improving the performance of the Support Vector Machine (SVM) and Random Forest classification methods for diabetes disease prediction. The methods used in this research were SVM and Random Forest classification methods with SMOTE-ENN. The SMOTE-ENN method was used to balance the diabetes data and remove noise data adjacent to the majority and minority classes. Data that has been balanced was predicted using SVM and Random Forest methods based on the division of training and testing data with 10-fold cross-validation. The results of this study were Random Forest method with SMOTE-ENN got the best performance compared to the SVM method, such as accuracy of 95.8%, sensitivity of 98.3%, and specificity of 92.5%. In addition, the proposed method approach (Random Forest with SMOTE-ENN) also obtained the best accuracy compared to previous studies referenced. Thus, the proposed method can be adopted to predict diabetes in a health application.

Keywords—SMOTE-ENN; data imbalance; SVM; random forest; health dataset

I. INTRODUCTION

Machine learning methods on health data, especially disease classification, have been widely practiced. The problem is that the dataset's quality influences the performance of machine learning methods in disease classification. In general, most health data, especially disease data, have data imbalance problems, such as diabetes [1], heart [2][3], and breast cancer [4]. If the problem of unbalanced data in health datasets is not addressed, it can affect the performance of classification methods, making the prediction results biased. With balanced data, classification methods can easily predict the majority class more accurately than the minority class. Therefore, this research seeks a method approach for handling unbalanced data on health data, especially diabetes disease data, so that classification methods achieve optimal accuracy.

Some previous studies have predicted diseases using various approaches, such as research [5] using the logistic regression machine learning method with SMOTE for predicting diabetes with an accuracy of 77%, precision of 75%, recall of 77%, and F1-score 76%. Research [6] uses forward

chaining and certainty factor methods to diagnose types of rheumatic diseases with an accuracy of 80%. Research [7] uses the SMOTE method approach with machine learning algorithms such as Xgboost, Random Forest, KNN, Logistic regression, Decision Tree, Naive Bayes, and XGBoost for liver disease prediction with an accuracy of 80%. Based on the results of their research, the XGBoost method with SMOTE produces better performance than other methods, with accuracy of 93%, Recall of 97%, Precision of 92%, and F1-Score of 94%.

Research [4] uses a hybrid sampling method (SMOTE and SpreadSupsample) with several machine learning methods such as Naive Bayes, Decision Tree C4.5, and Random Forest for breast cancer disease prediction. His research shows that the use of hybrid sampling can improve the performance of the machine learning methods used, such as accuracy, ROC, Recall, and Precision. Research [8] uses hybrid sampling (SMOTE-ENN) with the Artificial Neural Network (ANN) method for the identification of Marburg virus inhibitors. The results show that using hybrid sampling (SMOTE-ENN) can effectively increase the ANN method's accuracy. Research [9] uses a hybrid sampling approach (M-SMOTE-ENN) with the Random Forest calcification method to solve unbalanced data problems in health data. The results show that using hybrid sampling (M-SMOTE and ENN) can improve the performance of the Random Forest method better than oversampling SMOTE and ENN individually without being combined.

Research [10] uses machine learning methods such as KNN, Decision Tree, Naïve Bayes, Random Forest, SVM, and histogram-based gradient boosting (HBGB) for diabetes prediction. The results show that the HBGB method performs better than other methods, with an accuracy of 92%. Research [11] compares several classification methods in machine learning for diabetes detection. The results show that the XGBoost method has better accuracy than other models, which is 94%. Research [12] uses a combination feature selection approach with several machine learning classification methods for diabetes detection. The results show that feature selection methods can improve the classification methods' accuracy. Random Forest is the method that gets the best accuracy, with a feature selection of 80%.

Research [13] uses a hybrid sampling approach (SMOTE-Tomek Link) with the Random Forest method for predicting diabetes. At the same time, the results show that the hybrid sampling method (SMOTE-Tomek Link) increases the

accuracy of the Random Forest method compared to SMOTE and Tomek Link separately. Research [14] predicts the risk of diabetes using a hybrid sampling approach (SMOTE-Tomek Link) with the ANN method. The results show that using hybrid sampling SMOTE-Tomek Link is better than SMOTE alone, with an accuracy of 92%.

Based on previous research, a gap can be improved; that is, the accuracy obtained in predicting diabetes is not optimal, so it can still be increased. Based on research [13][14], the highest accuracy is 92% using a hybrid sampling SMOTE-Tomek link with ANN. Therefore, this study proposes a hybrid sampling SMOTE-ENN approach to improving performance, such as accuracy, sensitivity, and specificity in SVM and Random Forest classification methods. This research adopts the use of the SMOTE-ENN hybrid sampling method, as it performs better than SMOTE-Tomek Links [15][16].

The purpose of this study is the implementation of hybrid sampling SMOTE-ENN to increase the accuracy of the machine learning method in predicting unbalanced diabetes data. This study consists of an introduction structure, research method, results and discussion, and conclusion.

II. RESEARCH METHOD

This study has several stages shown in Fig. 1. The first stage is the collection of diabetes disease datasets obtained from the Uci Repository with a total dataset of 768 instances and ten attributes. Attributes owned by the Pima Indian Diabetes contain datasets such as Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, and Outcome (Class). The dataset used has a total of two class categories, namely positive diabetes and negative diabetes.

The second stage is data preprocessing which is useful for improving the quality of the dataset used, such as removing missing values, outliers, and unbalanced data in the data preprocessing phase using data sampling to balance the data in the diabetes class, where there is a smaller number of positive classes (minority classes) compared to negative classes (majority classes) so that it can affect the performance of the classification method. If unbalanced diabetes data is not handled, the classification method will find it easier to classify the majority (negative) class than the minority (positive) class. In other words, the classification method makes biased prediction results.

This research uses several data sampling methods such as SMOTE, ENN, and hybrid sampling SMOTE-ENN. The SMOTE-ENN method combines SMOTE oversampling and ENN undersampling. The way the SMOTE-ENN method works is to add artificial data to the minority class by interpolating the original data using SMOTE so that the resulting artificial data is balanced. After the data is balanced, samples from the majority class adjacent to the minority class are removed by undersampling ENN. The use of the SMOTE-ENN method can reduce data overfitting and noise. The method of SMOTE-ENN can be shown in Fig. 2.

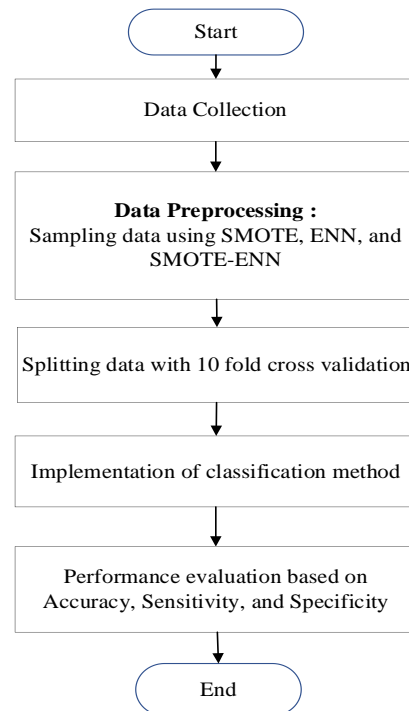


Fig. 1. Research stages.

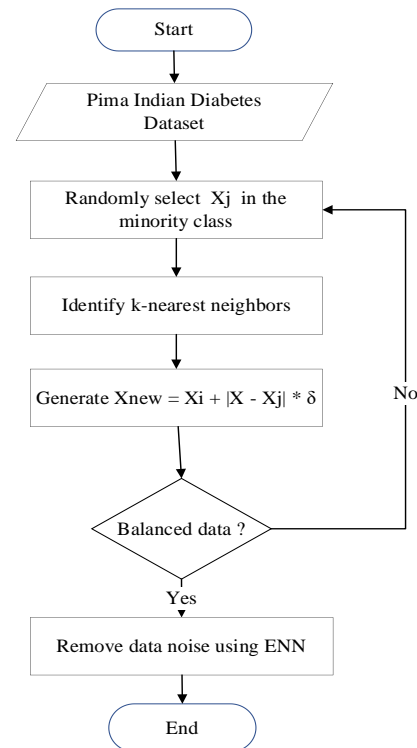


Fig. 2. SMOTE-ENN process.

After the data sampling, the next step is to divide the training and testing data using 10-fold cross-validation. 10-fold cross-validation works by dividing the data into ten groups, and each group can be used as training and testing data alternately. The illustration of how 10-fold cross-validation works is shown in Fig. 3.



Fig. 3. Process of 10-Fold cross validation.

Data divided into 10 folds are then used to implement classification methods using SVM and Random Forest methods. The classification results of the SVM and Random Forest methods are tested for performance based on accuracy, sensitivity, and specificity using the confusion matrix table. The accuracy, sensitivity, and specificity formulas use Equations (1), (2), and (3), respectively [17] [13].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

III. RESULT AND DISCUSSION

This section contains the results that have been achieved at each stage. The first stage is the collection of diabetes disease datasets obtained from the Uci Repository with a total of 768 instances and ten attributes. After data collection, the next step is data preprocessing. In the data preprocessing stage, it is used to improve data quality in diabetes disease data to optimize the classification method's performance.

There are unbalanced data in the data used so that it can reduce the performance of the classification method. The number of negative classes is 500 instances (majority class), and positive classes are 268 instances (minority class). This research proposes several sampling methods to balance the data: the SMOTE, ENN, and SMOTE-ENN hybrid. The amount of data generated by each sampling method is shown in Table I.

TABLE I. DATA DISTRIBUTION BEFORE AND AFTER SAMPLING

Sampling Method	Positive Class	Negative Class
Original Data	268	500
SMOTE	500	500
ENN	268	240
SMOTE-ENN	303	227

In Table I, the SMOTE method produces a balanced class by adding the minority class so that the number equals the majority class. However, the SMOTE method has the disadvantage of producing noise in the new data generated. The Edited Nearest Neighbor (ENN) method balances the data by removing the majority class (positive class) adjacent to the minority class so that it can reduce data noise in the dataset, while the SMOTE-ENN method makes the data balanced by

combining the SMOTE and ENN methods. The SMOTE method is used to add new data to the minority class based on the nearest neighbor. After the SMOTE results are balanced, the removal of adjacent data between the majority and minority classes is carried out to minimize data noise.

The data balanced using the sampling method is then divided into training and testing data using 10-fold cross-validation. Diabetes data is divided into training and testing, then implementing Random Forest and SVM classification methods for diabetes prediction. The classification results of the SVM and Random Forest methods are tested for performance based on accuracy, sensitivity, and specificity using the confusion matrix table. The confusion matrix results are obtained using the SVM method with original data (see Fig. 4), SMOTE result data (see Fig. 5), ENN method result data (see Fig. 6), and SMOTE-ENN data results (see Fig. 7).

Based on Fig. 4, the SVM method correctly classifies negative classes in as many as 438 instances, correctly classifies positive classes in as many as 151 instances, incorrectly classifies negative classes in as many as 62 instances, and incorrectly classifies positive classes in as many as 117 instances. The performance of the SVM method with the original data obtained an accuracy of 76.7%, a sensitivity of 56.3%, and a specificity of 87.6%.

Based on Fig. 5, the SVM method with SMOTE can correctly classify negative classes with 387 instances, correctly classify positive classes with 354 instances, incorrectly classify negative classes with 113 instances and incorrectly classify positive classes with 146 instances. The performance of the SVM method with SMOTE has an accuracy of 74.1%, sensitivity of 70.8%, and specificity of 77.4%.

Based on Fig. 6, the SVM method with ENN can correctly classify negative classes in as many as 207 instances, correctly classify positive classes in as many as 229 instances, incorrectly classify negative classes in as many as 33 instances and incorrectly classify positive classes in as many as 39 instances. The performance of the SVM method with ENN has an accuracy of 85.8%, sensitivity of 85.4%, and specificity of 86.3%.

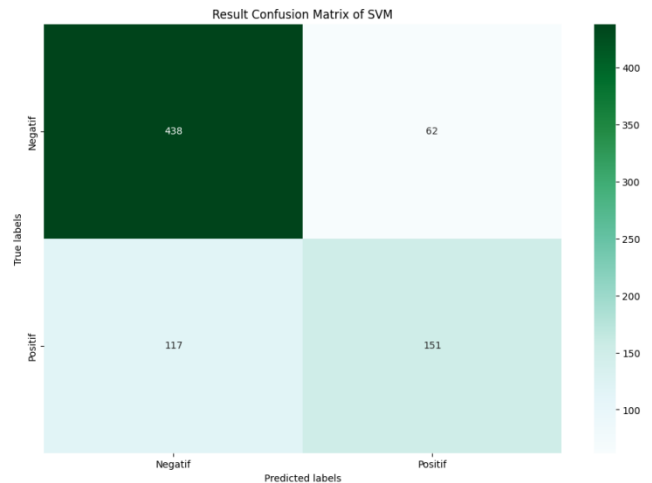


Fig. 4. SVM results with original data.

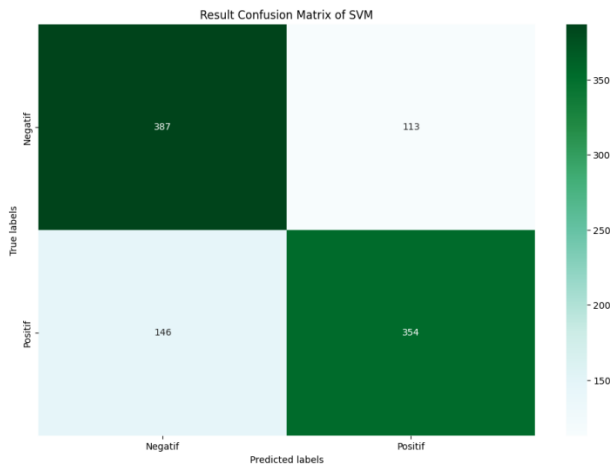


Fig. 5. SVM results with result data.

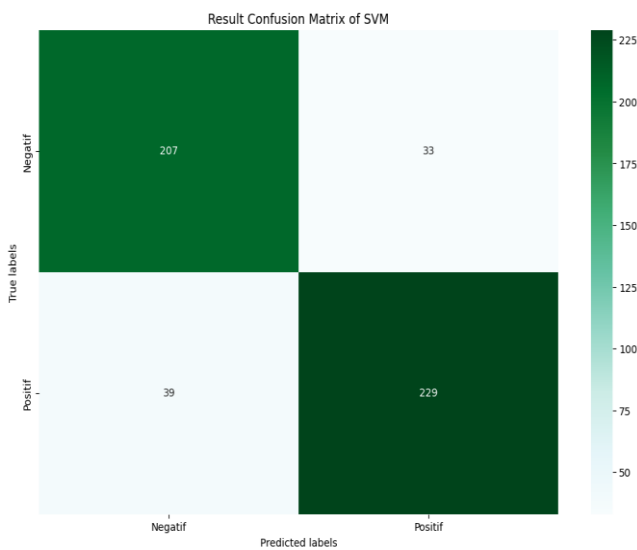


Fig. 6. SVM results with ENN result data.

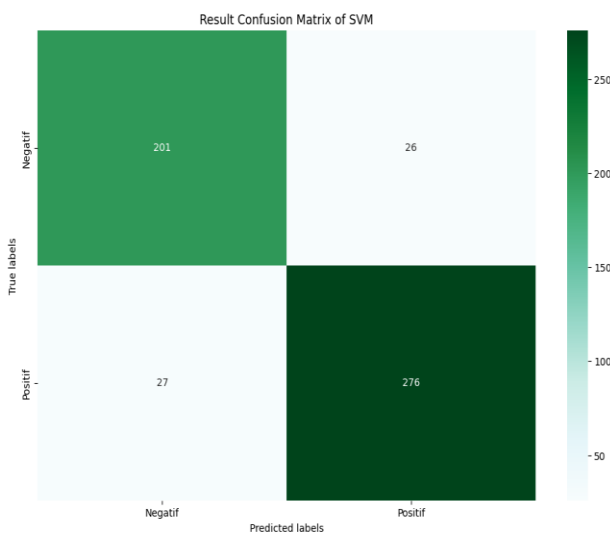


Fig. 7. SVM results with SMOTE-ENN result data.

Based on Fig. 7, the SVM method with SMOTE-ENN can classify negative classes correctly in as many as 201 instances, classify positive classes correctly in as many as 276 instances, classify negative classes incorrectly in as many as 26 instances, and classify positive classes incorrectly as many as 27 instances. The performance of the SVM method with SMOTE-ENN gets an accuracy of 90%, sensitivity of 91.1%, and specificity of 88.5%.

Then the confusion matrix results using the Random Forest method with original data (See Fig. 8), SMOTE data (See Fig. 9), ENN data (See Fig. 10), and SMOTE-ENN data (See Fig. 11).

Based on Fig. 8, the Random Forest method correctly classifies negative classes in as many as 429 instances, correctly classifies positive classes in as many as 156 instances, incorrectly classifies negative classes in as many as 72 instances, and incorrectly classifies positive classes in as many as 112 instances. The performance of the Random Forest method with the original data obtained an accuracy of 76.1%, sensitivity of 58.2%, and specificity of 85.8%.

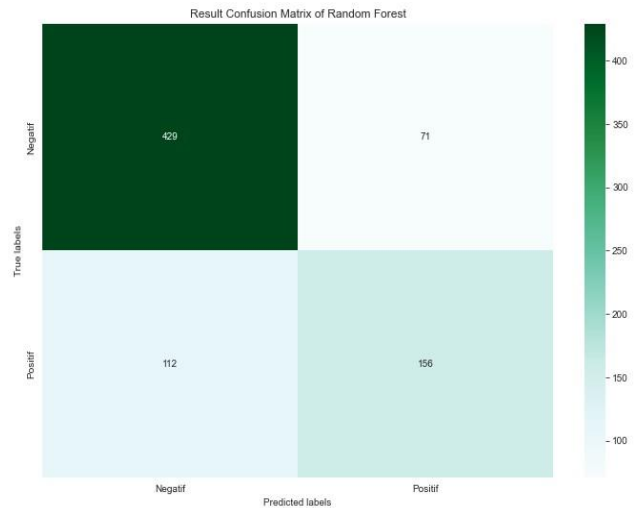


Fig. 8. Random forest results with original data.

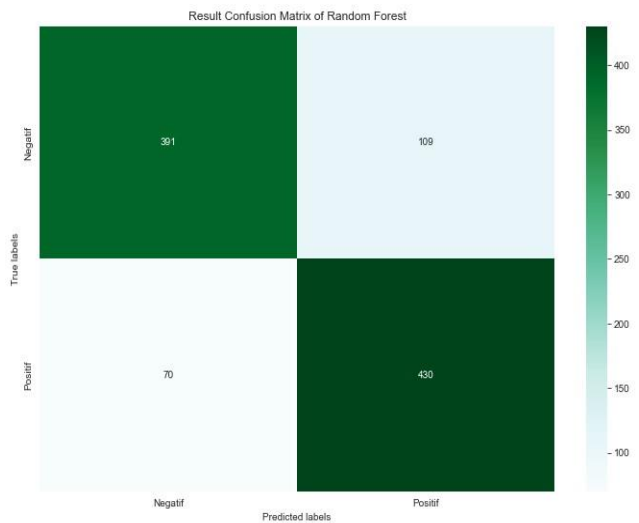


Fig. 9. Random forest results with SMOTE.

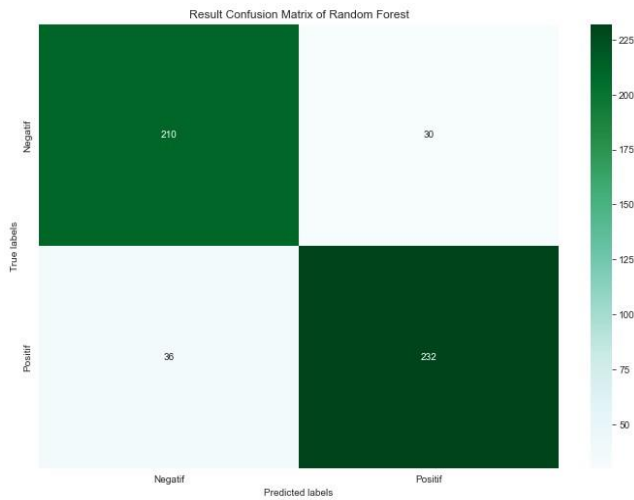


Fig. 10. Random forest results with ENN.

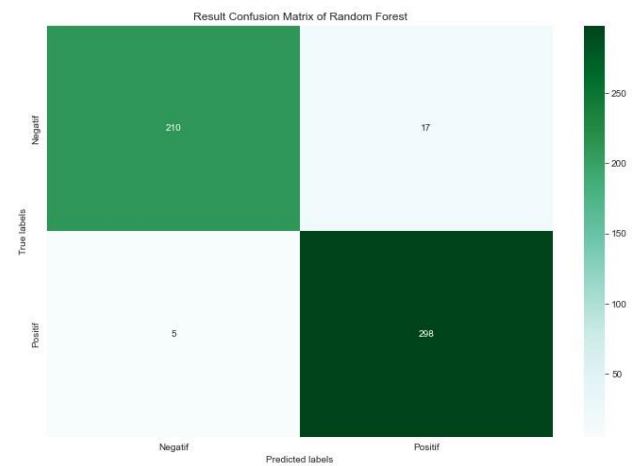


Fig. 11. Random forest results with SMOTE-ENN.

Based on Fig. 9, the Random Forest method with SMOTE can correctly classify negative classes in as many as 391 instances, correctly classify positive classes in as many as 430 instances, incorrectly classify negative classes in as many as 109 instances and incorrectly classify positive classes in as many as 70 instances. The performance of the Random Forest method with SMOTE gets an accuracy of 82.1%, sensitivity of 86%, and specificity of 78.2%.

Based on Fig. 10, the Random Forest method with ENN can correctly classify negative classes in as many as 210 instances, correctly classify positive classes in as many as 232 instances, incorrectly classify negative classes in as many as 30 instances and incorrectly classify positive classes in as many as 36 instances. The performance of the Random Forest method with ENN has an accuracy of 87%, sensitivity of 86.6%, and specificity of 87.5%.

Based on Fig. 11, the Random Forest method with SMOTE-ENN can correctly classify negative classes in as many as 210 instances, correctly classify positive classes in as many as 298 instances, incorrectly classify negative classes in as many as 17 instances and incorrectly classify positive classes as many as 5 instances. The performance of the

Random Forest method with SMOTE-ENN gets an accuracy of 95.8%, sensitivity of 98.3%, and specificity of 92.5%.

The results can be seen in Table II to simplify the understanding of the research results achieved based on several experiments that have been carried out.

TABLE II. CLASSIFICATION METHOD PERFORMANCE RESULTS WITH DATA SAMPLING APPROACH

	Accuracy	Sensitivity	Specificity
SVM	76,7%	56,3%	87,6
SVM with SMOTE	74,1%	70,8%	77,4%
SVM with ENN	85,8%	85,4%	86,3%
SVM with SMOTE-ENN	90%	91,1%	88,5%
Random Forest	76,1%	58,2%	85,8%
Random Forest with SMOTE	82,1%	86%	78,2%
Random Forest with ENN	87%	86,6%	87,5%
Random Forest with SMOTE-ENN	95,8%	98,3%	92,5%

Based on Table II, the Random Forest method with SMOTE-ENN produces the highest performance compared to SVM, with an accuracy of 95.8%, sensitivity of 98.3%, and specificity of 92.5%. Furthermore, the approach using the SMOTE-ENN sampling method resulted in better average performance than the SMOTE and ENN methods separately, such as accuracy, sensitivity, and specificity. The SMOTE-ENN sampling method is better than SMOTE and ENN separately because it can minimize noise data in the artificial data produced. The noise data in this context is the minority class data that is close to the majority class, so the classification method makes biased predictions. Besides that, using hybrid sampling by combining oversampling and undersampling methods in solving unbalanced data performs better than oversampling without undersampling [18]. SMOTE-ENN hybrid sampling in this study can significantly improve the sensitivity performance [19][20]. In order to see that the method proposed in this study is better than some related previous studies, the following comparison of the results can be seen in Table III.

TABLE III. RESULTS COMPARISON WITH PREVIOUS RESEARCH

Previous Studies	Methods	Scope of Study	Accuracy
Hairani et al. [13]	Random Forest and SMOTE-Tomek Links	Diabetes Disease	86,4%
ElSeddawy, et al. [14]	ANN + Gridsearch + SMOTE		92%
Sabhita et al. [21]	SVM + RFE + SMOTE		82%
Abdullah, et al. [22]	Random Forest + SMOTE		83%
Ijaz et al. [23]	DBSCAN + SMOTE + Random Forest		83,6%
Butt et al. [24]	LSTM		87,3%
Proposed Method	Random Forest and SMOTE-ENN		95,8%

IV. CONCLUSION

Based on the research results obtained using the SVM and Random Forest methods combined with the SMOTE, ENN, and hybrid SMOTE-ENN sampling methods, the Random Forest method with SMOTE-ENN produces better performance than the SVM method based on an accuracy of 95.8%, sensitivity of 98.3%, and specificity of 92.5% in diabetes prediction. Moreover, it can also be concluded that the SMOTE-ENN sampling method produces better performance than the SMOTE method without ENN in the results of the classification method used. Future researchers are suggested to use the ensemble learning method to improve the performance of the classification method.

ACKNOWLEDGMENT

Thanks to DRTPM of the Ministry of Education, Culture, Research, and Technology for funding under the Regular Fundamental Research scheme in 2023.

REFERENCES

- [1] H. Hairan, K. E. Saputro, and S. Fadli, "K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, pp. 89–93, Apr. 2020, doi: 10.14710/jtsiskom.8.2.2020.89-93.
- [2] E. Erlin, Y. Desnelita, N. Nasution, L. Suryati, and F. Zoromi, "Impact of SMOTE on Random Forest Classifier Performance based on Imbalanced Data," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 677–690, 2022, doi: 10.30812/matrik.v21i3.1726.
- [3] K. Wang *et al.*, "Improving risk identification of adverse outcomes in chronic heart failure using smote +enn and machine learning," *Risk Manag. Healthc. Policy*, vol. 14, no. May, pp. 2453–2463, 2021, doi: 10.2147/RMHP.S310295.
- [4] K. Rajendran, M. Jayabalan, and V. Thiruchelvam, "Predicting breast cancer via supervised machine learning methods on class imbalanced data," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 54–63, 2020, doi: 10.14569/IJACSA.2020.0110808.
- [5] Erlin, Y. N. Marlim, Junadhi, L. Suryati, and N. Agustina, "Early Detection of Diabetes Using Machine Learning with Logistic Regression Algorithm," *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022.
- [6] Hairani, M. N. Abdillah, and M. Innuddin, "An Expert System for Diagnosis of Rheumatic Disease Types Using Forward Chaining Inference and Certainty Factor Method," in *2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, 2019, pp. 104–109. doi: 10.1109/SIET48054.2019.8986035.
- [7] J. Yang and J. Guan, "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm," *Information*, vol. 13, no. 10, pp. 1–15, Oct. 2022, doi: 10.3390/info13100475.
- [8] M. Kumari and N. Subbarao, "A hybrid resampling algorithms SMOTE and ENN based deep learning models for identification of Marburg virus inhibitors," *Future Med. Chem.*, vol. 14, no. 10, pp. 701–715, Apr. 2022, doi: 10.4155/fmc-2021-0290.
- [9] Z. Xu, D. Shen, T. Nie, and Y. Kou, "A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data," *J. Biomed. Inform.*, vol. 107, no. June, p. 103465, 2020, doi: 10.1016/j.jbi.2020.103465.
- [10] R. Islam, A. Sultana, M. N. Tuhin, M. S. H. Saikat, and M. R. Islam, "Clinical Decision Support System for Diabetic Patients by Predicting Type 2 Diabetes Using Machine Learning Algorithms," *J. Healthc. Eng.*, vol. 2023, pp. 1–11, May 2023, doi: 10.1155/2023/6992441.
- [11] D. Sumathi, "Implementing a Model to Detect Diabetes Prediction using Machine Learning Implementing a Model to Detect Diabetes Prediction using Machine Learning Classifiers," *J. Algebr. Stat.*, vol. 13, no. 1, pp. 558–566, 2022.
- [12] R. Saxena, S. K. Sharma, M. Gupta, and G. C. Sampada, "A Novel Approach for Feature Selection and Classification of Diabetes Mellitus: Machine Learning Methods," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–11, Apr. 2022, doi: 10.1155/2022/3820360.
- [13] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using Smote-Tomek Link," *Int. J. Informatics Vis.*, vol. 7, no. 1, pp. 258–264, 2023.
- [14] A. I. ElSeddawy, F. K. Karim, A. M. Hussein, and D. S. Khafaga, "Predictive Analysis of Diabetes-Risk with Class Imbalance," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–16, Oct. 2022, doi: 10.1155/2022/3078025.
- [15] U. Ependi, A. F. Rochim, and A. Wibowo, "A Hybrid Sampling Approach for Improving the Classification of Imbalanced Data Using ROS and NCL Methods," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 3, pp. 345–361, 2023, doi: 10.22266/ijies2023.0630.28.
- [16] T. Sasada, Z. Liu, T. Baba, K. Hatano, and Y. Kimura, "A resampling method for imbalanced datasets considering noise and overlap," in *Procedia Computer Science*, 2020, vol. 176, pp. 420–429. doi: 10.1016/j.procs.2020.08.043.
- [17] H. Hairani, A. Anggrawan, A. I. Wathan, K. A. Latif, K. Marzuki, and M. Zulfikri, "The Abstract of Thesis Classifier by Using Naive Bayes Method," in *Proceedings - 2021 International Conference on Software Engineering and Computer Systems and 4th International Conference on Computational Science and Information Management, ICSECS-ICOCSIM 2021*, 2021, no. August, pp. 312–315. doi: 10.1109/ICSECS52883.2021.00063.
- [18] W. Nugraha, R. Maulana, Latifah, P. A. Rahayuningsih, and Nurmalasari, "Over-sampling strategies with data cleaning for handling imbalanced problems for diabetes prediction," *AIP Conf. Proc.*, vol. 2714, no. 1, p. 30017, May 2023, doi: 10.1063/5.0128407.
- [19] S. Balasubramanian, R. Kashyap, S. T. CVN, and M. Anuradha, "Hybrid Prediction Model For Type-2 Diabetes With Class Imbalance," in *2020 IEEE International Conference on Machine Learning and Applied Network Technologies (ICMLANT)*, 2020, pp. 1–6. doi: 10.1109/ICMLANT50963.2020.9355975.
- [20] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," *Int. J. Inf. Educ. Technol.*, vol. 13, no. 2, pp. 289–295, 2023, doi: 10.18178/ijiet.2023.13.2.1806.
- [21] E. Sabitha and M. Durgadevi, "Improving the Diabetes Diagnosis Prediction Rate Using Data Preprocessing, Data Augmentation and Recursive Feature Elimination Method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 921–930, 2022, doi: 10.14569/IJACSA.2022.01309107.
- [22] M. N. Abdullah and Y. B. Wah, "Improving Diabetes Mellitus Prediction with MICE and SMOTE for Imbalanced Data," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 2022, pp. 209–214. doi: 10.1109/AiDAS56890.2022.9918773.
- [23] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid Prediction Model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, Synthetic Minority Over Sampling Technique (SMOTE), and random forest," *Appl. Sci.*, vol. 8, no. 8, 2018, doi: 10.3390/app8081325.
- [24] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, "Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications," *J. Healthc. Eng.*, vol. 2021, pp. 1–17, Sep. 2021, doi: 10.1155/2021/9930985.

An Ensemble Load Balancing Algorithm to Process the Multiple Transactions Over Banking

Raghunadha Reddi Dornala
Cloud Architect, USA

Abstract—The banking industry has been transformed by cloud computing, which has provided scalable and cost-effective solutions for managing large volumes of transactions. However, as the number of transactions grow, the need for efficient load-balancing algorithms to ensure optimal utilization of cloud resources and improve system performance becomes critical. This paper proposes an ensemble cloud load-balancing (ECBA) algorithm specifically designed to process multiple banking transactions. The proposed algorithm combines the strengths of several load-balancing techniques to achieve a balanced distribution of transaction loads in various cloud servers. It considers factors such as transaction types, server capacities, and network conditions to make intelligent load distribution decisions. The algorithm dynamically adapts to changing workload patterns and optimizes resource allocation by leveraging machine learning and predictive analytics. A simulation environment that mimics the banking system's transaction processing workflow is created to evaluate the performance of the ensemble load balancing algorithm. Extensive experiments with various workload scenarios are conducted to assess the algorithm's effectiveness in load balancing, response time, resource utilization, and overall system performance. The results show that the proposed ECBA outperforms traditional banking load-balancing approaches. It reduces response time, improves resource utilization, and ensures every server is adequately funded with a few transactions. The algorithm's adaptability and scalability make it well-suited for handling dynamic and fluctuating workloads, thus providing a robust solution for processing multiple transactions in the banking sector.

Keywords—Cloud computing; load balancing; ensemble algorithm; banking; transaction processing; resource utilization; response time; scalability

I. INTRODUCTION

Load balancing techniques are critical for ensuring efficient resource utilization and maintaining optimal performance in cloud computing environments [1] [2]. Cloud computing provides users with vast computational resources and services. Still, the dynamic nature of cloud environments, the variability of workloads, and the scale of resources present several load-balancing challenges. This section investigates the issues and challenges associated with cloud computing load-balancing techniques. Discuss the complexities introduced by cloud systems' distributed nature, workload variability, scalability, and fault tolerance and the importance of ensuring fairness and resource utilization [3]. Understanding these difficulties is critical for developing effective load-balancing mechanisms in cloud environments. Cloud computing environments typically comprise several

geographically dispersed data centers or clusters [4]. Managing the load across these disparate resources is a difficult task. Load-balancing algorithms must consider network latency, bandwidth constraints, and the availability of resources across multiple locations. Coordination of workload distribution and efficient resource utilization in such environments is a significant challenge [5].

Because of the dynamic nature of user demands, cloud systems experience highly variable workloads. The load on cloud resources can fluctuate dramatically, necessitating real-time load-balancing techniques. Predicting and managing these workload variations is critical to avoid resource underutilization or overload situations. Load-balancing algorithms must consider historical and real-time workload data to make intelligent workload distribution decisions [6]. Cloud computing is intended to scale horizontally, allowing for adding or removing resources in response to demand. Load balancing techniques must be able to adjust workload distribution as resources scale up or down dynamically [7].

Furthermore, cloud systems are vulnerable to failures and faults. Load balancers should be fault-tolerant, detecting and redirecting workloads away from failed or degraded resources to ensure high availability and reliability. Load-balancing algorithms should strive for equity and fair resource distribution among users or applications. Maintaining user satisfaction and avoiding performance bottlenecks requires ensuring each user or application receives a fair share of resources. Balancing workload distribution while considering workload priorities, user requirements, and resource constraints, can be difficult.

In the fast-paced world of banking and finance, effective transaction management is critical for ensuring customer satisfaction, operational stability, and data security. Traditional load-balancing algorithms frequently fall short of distributing workloads evenly across resources, owing to the increasing reliance on digital transactions and the ever-increasing volume of data. This paper proposes an innovative ensemble load-balancing algorithm designed to optimize banking transaction processing. The proposed algorithm aims to improve system performance, ensure resource utilization, and provide a seamless customer experience by combining the strengths of multiple load-balancing techniques. The banking sector is critical to global economic activity, processing a wide range of transactions daily. As more customers use digital banking services, there is a greater need for seamless and quick transaction processing. Banks rely heavily on information technology systems and networks to meet these

demands, making efficient load balancing critical for ensuring optimal performance.

The organization of this paper is as follows. Section II explains various cloud models applied on multiple datasets to analyze the performance. Section III describes the weighted Round Robin algorithm in the banking sector. Section IV presents the adaptive load balancing and its functionalities. The fifth section explains the proposed combined approach and its functionalities. Section VI gives the performance metrics used in this paper. Section VII shows the comparative performances of various existing and proposed algorithms. The final section provides the conclusion of the overall research work.

II. LITERATURE SURVEY

X. Wei et al. [8] introduced the popularity-based position technique that maps data components and edge servers to retrieve the virtual coordinate in the plane. The proposed model performance is improved to tackle the load balancing between edge servers via offloading. Experiments show that the proposed model effectively reduces the average path length for data access, and load-balancing techniques provide better options for overloaded servers. T. Liu et al. [9] introduced the novel Q-networks that will allocate resources using resource computation. The main objective of novel q-networks used to decrease the latency over a long time. The performance of Novel q-networks shows better performance in terms of performance. S. Nath et al. [10] proposed an automated scheduling model that incorporates Deep Reinforcement Learning (DRL) and the Deep Deterministic Policy Gradient (DDPG) method. The DDPG extracts optimized models to develop multi-cell MEC systems by leveraging cooperation among neighboring MEC servers. When compared to DDPG, the existing model produced better results. J. Zhang et al. [11] discussed several comparative performances based on the security and privacy threats that edge computing can address. This work primarily focused on discussing various issues and factors that aid in developing several edge computing applications. The study also provides solutions for several privacy and security issues based on edge-related paradigms. T. Li et al. [12] introduced privacy grouping issues that reduce problems in edge clouds and thus reduce edge cloud maintenance. The grouping techniques carefully developed the optimized goal for two models such as tree-based hierarchical (TBH) and graph-based interconnected (GBI) edge clouds. The proposed model obtained better outcomes on two benchmark data sets. B. Pourghebleh et al. [13] introduced the meta-heuristic model that solves the VM unification issue compared with the existing models based on the influential factors. E. H. Houssein et al. [14] introduced a different model belongs to task scheduling that classifies the cloud applications based on the scheduling issues which is one or multiple objectives. S. K. Mishra et al. [15] proposed an iterative approach that optimized the metrics such as latency and energy consumption and unloading the work and associated VM for the execution of the task. If the particular edge center is not ready to provide the resources, the user's request will send to the other cloud system. B. Alankar et al. [16] developed a combined approach that solves various issues in load balancing based on HAProxy, clusters in VM, and

cloud servers. Results show that the proposed model obtained better performance regarding load balancing metrics. A. A. Abdellatif et al. [17] proposed the SDN-based load balancing that reduces the high usage of resources and decreases the computation time. The proposed model executes the program on top of the SDN model and controls the tasks. The manager in this model maintains the transmission messages, maintains hosting pools, and checks the load status at peak time. M. A. Mukweho et al. [18] presented a comparative survey of work on fault tolerance applications used in the cloud environment. A better future model is required to improve the proposed approach's performance. B. Cao et al. [19] detected the replacement issue by using the edge servers (ESs) in the IoV to design various objectives used to deploy the applications and measure the task loading, energy usage, deployment expenses, etc. The proposed model also solves the ES deployment issue. B. Lin et al. [20] introduced the GA-DPSO combined with the genetic approach that optimizes data transmission based on the proposed flow. GA-DPSO is the combined domains such as edge computing and cloud computing. S. Yang et al. [21] proposed that the issue belongs to VNFs on the combined platform and analyzed the VM traffic present in the VMs. It is an integrated approach that uses various technologies that help improve the detection of abnormal traffic in cloud servers and reduce the traffic by controlling the user's requests. J. Zhang et al. [22] introduced the model deployed in a cloud server with an advanced domain called vehicular networks. Several high-potential models, such as fiber-wireless (FiWi), improve the vehicular edge computing networks (VECNs), analyze the task loading, and measure the vehicle's delay. The proposed model obtained superior results compared with existing approaches.

III. WEIGHTED ROUND ROBIN ALGORITHM FOR BANKING TRANSACTIONS

Maintaining unlimited banking transactions requires practical load-balancing algorithms that guarantee the network can handle many transactions without complex resources. The Weighted Round Robin (WRR) algorithm is a scheduling technique commonly used in banking systems to allocate resources for processing transactions. It aims to provide fairness and efficient utilization of resources based on predefined weights assigned to each transaction type. Here's an explanation of the WRR algorithm using equations:

A. Define Inputs

N: Number of transaction types.

W[i]: Weight assigned to each transaction type i, where i ranges from 1 to N.

Calculate the Weighted Round Robin Quantum (Q):

Calculate the total weight sum (TW) as the sum of all transaction weights:

$$TW = W[1] + W[2] + \dots + W[N] \quad (1)$$

Set the Quantum (Q) as the least common multiple (LCM) of all transaction weights:

$$Q = \text{LCM}(W[1] + W[2] + \dots + W[N]) \quad (2)$$

B. Initialize Variables

Current_quantum: The current quantum being processed initially set to 0.

Current_transaction_type: The index of the currently selected transaction type initially set to 0.

C. Transaction Scheduling

- Iterate through the transactions in a loop:
- For each transaction, perform the following steps:
 - Increment the current_quantum by 1.
 - If current_quantum is equal to or exceeds the Quantum (Q), perform the following steps:
 - Reset current_quantum to 0.
 - Move to the next transaction type by incrementing current_transaction_type by 1.
 - If current_transaction_type exceeds the maximum index (N), set current_transaction_type back to 1.
 - Select the transaction type indicated by current_transaction_type for processing.

Output: The output of the WRR algorithm is the sequence of transaction types selected for processing, based on their assigned weights and the calculated quantum.

The Weighted Round Robin algorithm ensures that transactions with higher weights receive a proportionally higher share of the available processing resources. By defining appropriate weights for each transaction type, the algorithm can be customized to prioritize certain types of transactions or balance the workload across different transaction types.

IV. ADAPTIVE LOAD BALANCING (ALB) FOR BANKING TRANSACTIONS IN CLOUD COMPUTING

ALB mainly focuses on transactions made by the users in the cloud server, which involves several distributed workload among multiple servers to advance performance and shows effective transactions. Here are the steps and equations involved in adaptive load balancing:

1) *Measure server performance:* The first step is to collect data on the performance of each server in the cloud environment. This can include metrics such as CPU utilization, memory usage, network bandwidth, and response time.

2) *Determine the load balancing algorithm:* Choose an appropriate load balancing algorithm based on the specific requirements of banking transactions. Some commonly used algorithms include round-robin, least connections, weighted round-robin, and least response time.

3) *Define the criteria for load balancing:* Establish the criteria for load balancing decisions. This can include thresholds for server utilization, response time, or other performance metrics. For example, if a server's CPU utilization exceeds a certain threshold, it may be considered overloaded.

4) *Calculate server weights:* If using a weighted load balancing algorithm, assign weights to each server based on its capacity and performance characteristics. This allows for more fine-grained control over the distribution of the workload.

5) *Monitor server performance:* Continuously monitor the performance of each server in the cloud environment. This can be done using monitoring tools or by collecting real-time performance metrics.

6) *Evaluate server conditions:* Compare the performance metrics of each server against the defined criteria for load balancing. If a server exceeds the thresholds or is underutilized, it may be a candidate for load redistribution.

7) *Calculate the load factor:* Calculate the load factor for each server based on its current workload and capacity. The load factor can be calculated using various equations, such as:

$$\text{Load Factor} = \frac{\text{Current Server Load}}{\text{Maximum Server Capacity}} \quad (3)$$

This equation gives a normalized value between 0 and 1, representing the relative load on each server.

8) *Make load balancing decisions:* Based on the load factors and the chosen load balancing algorithm, make decisions on redistributing the workload. This involves determining which server should receive new requests or reassigning existing requests from overloaded servers to underutilized ones.

9) *Redirect requests:* Implement the load balancing decisions by redirecting incoming requests to the selected servers. This can be achieved through DNS-based load balancing, where the DNS server maps the domain name to the IP address of the appropriate server.

10) *Monitor and adjust:* Continuously monitor the system performance and reevaluate load balancing decisions as the workload and server conditions change. Adjust the load balancing parameters, such as thresholds or weights, if necessary, to further optimize performance.

V. WEIGHTED ROUND ROBIN ALGORITHM COMBINED WITH ADAPTIVE LOAD BALANCING IN CLOUD BANKING APPLICATION

Cloud computing has transformed how businesses operate in today's digital age, and the banking sector is no exception. Cloud banking applications take advantage of cloud infrastructure's scalability, flexibility, and cost-effectiveness to provide customers with efficient and dependable services. However, as the number of users and transactions grow, ensuring optimal performance and load balancing becomes increasingly essential for the smooth operation of these applications. The WRR algorithm and Adaptive Load Balancing techniques can be used to address this challenge. This effective combination provides improved resource allocation, faster response times, and more efficient use of cloud resources.

WRR is a load-balancing algorithm that cyclically distributes incoming requests across multiple servers. Each server is given a weight that corresponds to its processing

capacity. The algorithm considers these weights during load-balancing and distributes requests proportionally to each server. The WRR algorithm ensures that the workload is distributed relatively by assigning higher weights to more powerful servers, preventing any individual server from becoming overwhelmed.

The WRR algorithm is supplemented by Adaptive Load Balancing, which continuously monitors system performance metrics such as CPU utilization, memory usage, network traffic, and response times. The load balancer dynamically adjusts the weights assigned to each server based on these metrics. If a server begins to experience increased workloads or performance degradation, the load balancer can reduce its weight, redirecting traffic to other servers with lower workloads. If a server's performance or load improves, its importance can be increased, allowing it to handle more requests (see Fig. 1).

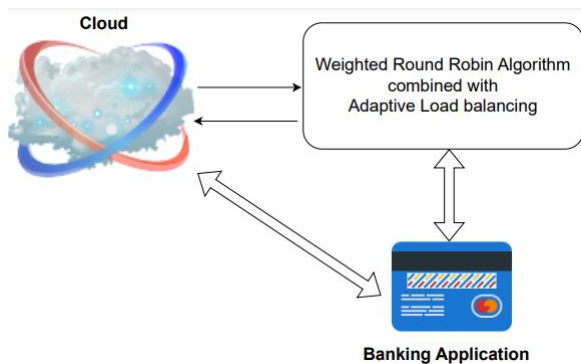


Fig. 1. Proposed architecture.

Combining the WRR algorithm with Adaptive Load Balancing in a cloud banking application provides numerous benefits. For starters, it ensures that each server is used to its full potential, avoiding bottlenecks and improving overall system performance. Second, it enables the application to adapt to changing traffic patterns and allocate resources dynamically based on real-time conditions, resulting in faster response times and a better user experience. Finally, it improves the application's scalability by allowing additional servers to be easily added or removed from the pool without disrupting the load-balancing mechanism.

VI. PERFORMANCE METRICS

Load balancing is a critical technique in cloud computing to distribute workloads across multiple servers or resources to optimize performance and ensure efficient resource utilization. The effectiveness of load-balancing algorithms is assessed using a variety of performance metrics. Here is some common load balancing performance metrics along with their equations:

1) *Response Time (RT)*: It measures the time taken by a server to respond to a client request. Lower response time indicates better performance.

$$RT = (T_f - T_s) + T_w$$

Where T_f the time of the last response is received, T_s is the time of the first request sent, and T_w is the waiting time in the queue.

Example: If a client sends a request at time $T_s = 10s$, the last response is received at time $T_f = 15s$, and the waiting time in the queue is $T_w = 2s$, then the response time would be $RT = (15 - 10) + 2 = 7s$.

2) *Throughput (TH)*: It represents the number of requests processed per unit of time. Higher throughput indicates better performance.

$$TH = \frac{\text{Total requests processed}}{\text{Time period}}$$

Example: If a server processes 1000 requests in 10 minutes, then the throughput would be $TH = 1000 / (10 * 60) = 1.67$ requests/s.

3) *Utilization (U)*: It measures the extent to which a server or resource is utilized. It is often represented as a percentage. Higher utilization indicates efficient resource usage.

$$U = \frac{\text{Total busy time}}{\text{Total time}} \times 100$$

Example: If a server is busy for eight hours (28,800 seconds) out of a total of 24 hours (86,400 seconds), then the utilization would be $U = (28,800 / 86,400) * 100 = 33.33\%$.

4) *Queue length (QL)*: It represents the number of requests waiting in the queue for processing. Lower queue length indicates better performance.

$$QL = \lambda * W$$

Where λ is the arrival rate of requests (requests per unit of time) and W is the average waiting time in the queue.

Example: If the arrival rate is $\lambda = 10$ requests per second and the average waiting time in the queue is $W = 5$ seconds, then the queue length would be $QL = 10 * 5 = 50$ requests.

5) *Server load (SL)*: It measures the amount of work being processed by a server or resource. It is often represented as a percentage. Lower server load indicates better performance.

$$SL = \frac{C}{T} \times 100$$

Where C is the average number of requests being processed by the server and T is the total capacity of the server (maximum number of requests it can handle in a given time period).

Example: If the average number of requests being processed by the server is $C = 30$ and the total capacity of the server is $T = 50$, then the server load would be $SL = (30 / 50) * 100 = 60\%$.

VII. EXPERIMENTAL RESULTS

Experiments were conducted using the cloud banking application developed using Python. The proposed integrated load balancing is implemented at the client and server sides. The comparative performance shows that the proposed model obtained better results. The performance of existing and proposed models was observed on 1k, 5k, and 10k transactions at a time. The existing model's round-robin (RR), Least Connection (LC), and Adaptive Load Balancing (ALB) compared with ECBA.

TABLE I. COMPARATIVE PERFORMANCES IN TERMS OF FOLLOWING METRICS FOR 1K TRANSACTIONS

Metrics	RR	LC	ALB	ECBA
Response time (Sec)	22	20	18	15.1
TH(R/S)	16.7	18.9	20.1	22.9
U (%)	70.7	72.9	75.3	80.3
QL (%)	60	55.6	52.1	49.1
SL (%)	70.8	66.7	63.1	59.3

In banking sector, daily multiple number of transactions will take place online. To process the large transactions an ECBA model obtain high results for 1k transactions given in Table I and graph shown in Fig. 2.

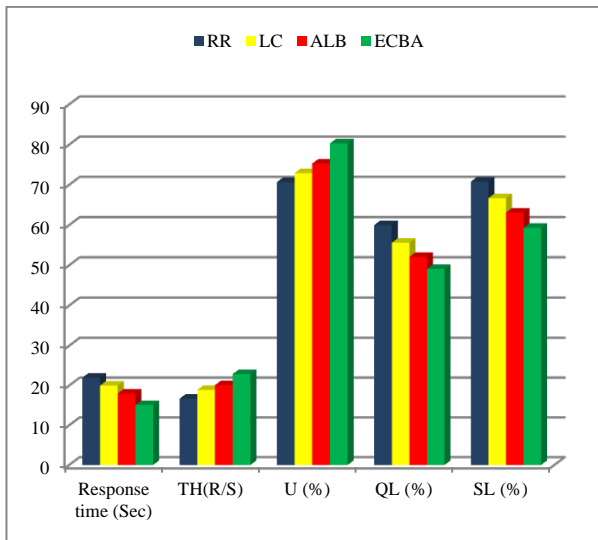


Fig. 2. Comparative performances in terms of following metrics for 1k transactions.

TABLE II. COMPARATIVE PERFORMANCES IN TERMS OF FOLLOWING METRICS FOR 5K TRANSACTIONS

Metrics	RR	LC	ALB	ECBA
Response time (Sec)	34	31.2	28	25.1
TH(R/S)	27.7	29.9	33.1	36.9
U (%)	75.7	77.4	79.2	83.3
QL (%)	69	65.6	61.1	55.1
SL (%)	80.8	66.7	63.1	59.3

Table II and Fig. 3 show the comparative performances of existing and proposed approaches based on given parameters. The overall transactions analyzed by the model are 5k. The high performance is achieved by ECBA and low performance is achieved by RR.

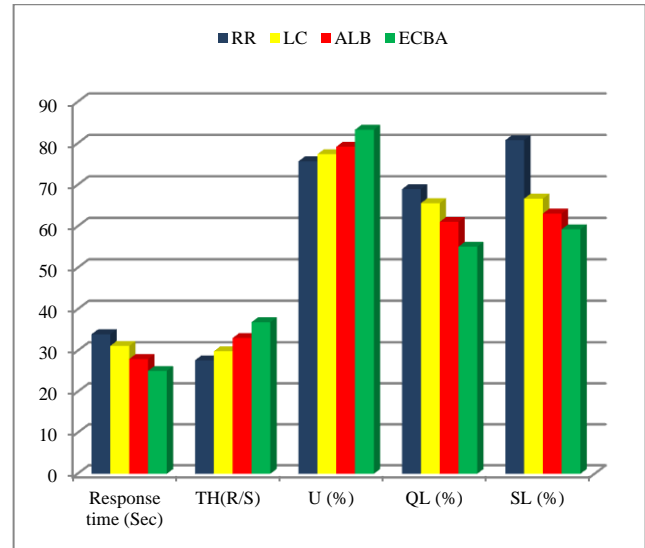


Fig. 3. Comparative performances in terms of following metrics for 5k transactions.

VIII. CONCLUSION

In ECBA, the WRR can balance the workload among multiple servers or instances in a cloud banking application to ensure efficient handling of customer requests. The server weights can assign memory capacity and network bandwidth based on the processing power. WRR can optimize resource utilization and prevent server overloading by allocating requests proportionally to the weights of the servers. Another technique used in cloud environments is adaptive load balancing (ALB), which dynamically adjusts the load-balancing algorithm based on real-time conditions. It continuously monitors server and network traffic performance and makes modifications to guarantee the best possible utilization of resources. ALB can assist in automatically modifying the load balancing algorithm in the context of a cloud banking application based on factors such as server response times, server availability, or network congestion. For example, suppose a server has long response times or is temporarily unavailable. In that case, ALB can route incoming requests to other available servers with lower loads, ensuring consistent performance and minimizing service disruptions. The combination of WRR and ALB can improve a cloud banking application's performance, scalability, and fault tolerance, ensuring efficient resource utilization and optimal customer experience. It enables the application to handle varying workloads while maintaining high availability, allowing dynamic adaptation to changing conditions. In the future, an ensemble load-balancing approach combined with security would be developed to achieve better performance.

REFERENCES

- [1] Kumar, P.; Kumar, R. Issues and challenges of load balancing techniques in cloud computing: A survey. *ACM Comput. Surv. (CSUR)* 2019, 51, 1–35.
- [2] Princess, G.A.P.; Radhamani, A. A Hybrid Meta-Heuristic for Optimal Load Balancing in Cloud Computing. *J. Grid Comput.* 2021, 19, 1–22.
- [3] Elmagzoub, M.A.; Syed, D.; Shaikh, A.; Islam, N.; Alghamdi, A.; Rizwan, S. A Survey of Swarm Intelligence Based Load Balancing Techniques in Cloud Computing Environment. *Electronics* 2021, 10, 2718. <https://doi.org/10.3390/electronics10212718>.
- [4] Mishra, S.K.; Sahoo, B.; Parida, P.P. Load balancing in cloud computing: A big picture. *J. King Saud Univ.-Comput. Inf. Sci.* 2020, 32, 149–158.
- [5] Junaid, M.; Sohail, A.; Ahmed, A.; Baz, A.; Khan, I.A.; Alhakami, H. A hybrid model for load balancing in cloud using file type formatting. *IEEE Access* 2020, 8, 118135–118155.
- [6] Shahid, M.A.; Islam, N.; Alam, M.M.; Mazliham, M.S.; Musa, S. Towards Resilient Method: An exhaustive survey of fault tolerance methods in the cloud computing environment. *Comput. Sci. Rev.* 2021, 40, 100398.
- [7] A. Kishor, R. Niyogi, A. T. Chronopoulos and A. Y. Zomaya, "Latency and Energy-Aware Load Balancing in Cloud Data Centers: A Bargaining Game Based Approach," in *IEEE Transactions on Cloud Computing*, vol. 11, no. 1, pp. 927-941, 1 Jan.-March 2023, doi: 10.1109/TCC.2021.3121481.
- [8] X. Wei and Y. Wang, "Popularity-Based Data Placement With Load Balancing in Edge Computing," in *IEEE Transactions on Cloud Computing*, vol. 11, no. 1, pp. 397-411, 1 Jan.-March 2023, doi: 10.1109/TCC.2021.3096467.
- [9] T. Liu, S. Ni, X. Li, Y. Zhu, L. Kong and Y. Yang, "Deep Reinforcement Learning Based Approach for Online Service Placement and Computation Resource Allocation in Edge Computing," in *IEEE Transactions on Mobile Computing*, vol. 22, no. 7, pp. 3870-3881, 1 July 2023, doi: 10.1109/TMC.2022.3148254.
- [10] S. Nath and J. Wu, "Deep reinforcement learning for dynamic computation offloading and resource allocation in cache-assisted mobile edge computing systems", *Intell. Converged Netw.*, vol. 1, no. 2, pp. 181-198, 2020.
- [11] J. Zhang, B. Chen, Y. Zhao, X. Cheng and F. Hu, "Data security and privacy-preserving in edge computing paradigm: Survey and open issues", *IEEE Access*, vol. 6, pp. 18209-18237, 2018.
- [12] T. Li et al., "Privacy-preserving participant grouping for mobile social sensing over edge clouds", *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 865-880, 2021.
- [13] B. Pourghebleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Comput.*, vol. 24, pp. 2673–2696, May 2021.
- [14] E. H. Houssein, A. G. Gad, Y. M. Wazery, and P. N. Suganthan, "Task scheduling in cloud computing based on meta-heuristics: Review, taxonomy, open challenges, and future trends," *Swarm Evol. Comput.*, vol. 62, Apr. 2021, Art. no. 100841.
- [15] S. K. Mishra, D. Puthal, B. Sahoo, S. Sharma, Z. Xue, and A. Y. Zomaya, "Energy-efficient deployment of edge datacenters for mobile clouds in sustainable IoT," *IEEE Access*, vol. 6, pp. 56587–56597, 2018.
- [16] B. Alankar, G. Sharma, H. Kaur, R. Valverde, and V. Chang, "Experimental setup for investigating the efficient load balancing algorithms on virtual cloud," *Sensors*, vol. 20, no. 24, p. 7342, Dec. 2020.
- [17] A. A. Abdellatif, E. Ahmed, A. T. Fong, A. Gani, and M. Imran, "SDN-based load balancing service for cloud servers," *IEEE Commun.Mag.*, vol. 56, no. 8, pp. 106–111, Aug. 2018.
- [18] M. A. Mukwevho and T. Celik, "Toward a smart cloud: A review of fault-tolerance methods in cloud systems," *IEEE Trans. Services Comput.*, vol. 14, no. 2, pp. 589–605, Mar. 2021.
- [19] B. Cao, S. Fan, J. Zhao, S. Tian, Z. Zheng, Y. Yan, and P. Yang, "Largescale many-objective deployment optimization of edge servers," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3841–3849, Jun. 2021.
- [20] B. Lin et al., "A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing", *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4254-4265, Jul. 2019.
- [21] S. Yang, F. Li, S. Trajanovski, X. Chen, Y. Wang and X. Fu, "Delay-aware virtual network function placement and routing in edge clouds", *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 445-459, Feb. 2021.
- [22] J. Zhang, H. Guo, J. Liu and Y. Zhang, "Task offloading in vehicular edge computing networks: A load-balancing solution", *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 2092-2104, Feb. 2020.

Genetic Approach for Improved Prediction of Adaptive Learning Activities in Intelligent Tutoring System

Fatima-Zohra Hibbi¹, Otman Abdoun², El Khatir Haimoudi³

SMARTiLAB, EMSI Rabat, Morocco¹

Advanced Science and Technologies Laboratory, Polydisciplinary Faculty,

Abdelmalek Essaadi University, Larache, Morocco^{1, 2, 3}

Abstract—The intelligent tutoring system registers the reference data of the learners in a database. This data is stored for later use in the instructional module. Designing a student model is not an easy task. It is first necessary to identify the knowledge acquired by the learner, then identify the learner's level of understanding of the functionality and finally identify the pedagogical strategies used by the learner to solve a problem. These elements must be taken into account in the development of the learner model. Learner characteristics must be considered in several forms. To build an effective learner model, the system must take into consideration both static (Learner preferences) and dynamic (Compartmental action) student characteristics. The objective of the article is to work out the learner model of the intelligent tutoring system by suggesting a new learning path. This proposal is based on the constructivist approach and the activist style (based on experimentation). According to the KOLB model, the authors propose a list of pedagogical activities depending on the learners' profile. Based on the learners' actions, the system reduces the list of activities based on two criteria: the learner's preference and the presence of one or more activities based on the activist style using genetic algorithm as an evolutionary algorithm. The results obtained led us to improve the learning process through a new conception of the ITS learner model.

Keywords—Intelligent tutoring system; learner model; genetic algorithm; adaptive learning activities

I. INTRODUCTION

e-Learning systems include smart devices for analysing, evaluating and assessing users' knowledge and skills, as well as monitoring and supervising the e-learning process. AI allows using and implementing its techniques to be more efficient educational systems [1], such as genetic algorithms, intelligent tutoring systems and neural networks. However, the problem that online learning systems lack is to optimize the learner activities during a learning process. Meanwhile, various researchers have already addressed this topic by providing several solutions [2], such as eye-tracking technology [3] clustering [4] and classification methods [5] to detect the learner style, etc. These solutions allow the detection of the learning style and subsequently deduce the list of appropriate activities.

Through experiential learning theory, the KOLB model has been the theoretical foundation for several models of learning

styles. Among these are the Dunn and Dunn [6] and Felder [7] models that developed the Learning Styles Inventory (LSI) to determine learning styles. This questionnaire consists of a self-evaluation questionnaire with approximately 100 questions that learners must answer. These questions are related to five categories. Several versions of this scale have been developed for adults. However, Felder and Silverman do not consider that all learners fit into predefined categories. For example, a learner may have the active-intuitive-global style; but have a strong preference for the active type, a weak preference for the intuitive and global brand and a moderate preference for the visual style. In this case, because he has a strong preference for the active type, he will likely have "great difficulty" learning in an environment that does not support this style. In the case of moderate visual style preference, he should learn "more easily" in a teaching environment that encourages visual style characteristics. For the intuitive and global styles, it is clear that the learner is well balanced on the scales of these two styles, so they would have no problem learning in an environment that favours one or both types simultaneously [8].

Categorizing learners according to their preferred learning style to match a set of complementary learning activities to each type is a promising idea. The preferred learner style affects the conception and construction of pedagogical activities as well as decision-making about the selection of learning resources [9]. From a pedagogical perspective, the teacher must be able to define a set of activities that correspond to each learning style, adapt the appropriate activity to an identified learning need, and design the most appropriate resources and activities for each individual according to their learning style. It should also be possible for a learner to examine the learning resources and activities associated with a particular learning session and assess whether they are appropriate for their preferred learning style.

The literature review allowed us to justify the uses of the KOLB model during this contribution. The main motivations for this choice are as follows:

A study of four learning style models and the experience of engineering educators in their practical applications are presented in [7].

The finding indicates that KOLB's model LIS helps learners learn the course because they have become aware of

their thought processes and helps them develop interpersonal skills.

The objective of this paper is to update the learner model of the intelligent tutoring system in such a way that the activist style will be a required style for the learning process. For this purpose, the authors suggest optimizing the list of activities proposed by the system by strengthening the practice-based exercises to improve the skills acquired by the learner using the evolutionary algorithm: genetic algorithm.

The paper is organized as follow: the second section presents a literature review of the learner module and its triggered problems. The upcoming section describes the choice and the implementation of the genetic algorithm. The fourth section illustrates and interprets the obtained results. Finally, this article is concluded with a summary of the contribution and the future scope.

II. LEARNER MODULE

It would be difficult for an intelligent tutoring system to succeed without a good understanding of the learner. The learner model represents the learner's knowledge and skills dynamically. Just as domain knowledge must be represented explicitly to be communicated, the learner model must also be represented in the same way. In principle, the learner model should store aspects of the learner's behaviour and skills so that the ITS can infer the learner's performance and skills [9].

The intelligent tutoring system keeps the learners' data in a database. It stores learner reference data such as name, ID, current level, overall score, course exercises completed, exercises difficulty level achieved, and several questions completed from a course exercise. This data is stored for later use in other modules such as the pedagogical module.

Designing a student model is not an easy task. It must be based on answers to specific questions. What does the learner know? What kind of knowledge would the learner need to solve a problem? The methodology for designing a learner model should be based on those questions. First, it is necessary to identify the knowledge acquired by the learner in terms of the components integrated into the mechanism.

Secondly, it is necessary to identify the learner's level of understanding of the functionality of the mechanism. Finally, it is required to determine the pedagogical strategies used by the learner to solve a problem. Those elements must be taken into account in the development of the learner model [10].

Learner characteristics must be considered in several forms. The system must take into account both static and dynamic features of the learner to construct an effective learner model. Static features include information such as email address, age, and native language and are defined before the learning processes begin. However, dynamic features come from the learners' behaviour during interaction with the system [11]. The Fig. 1 shows the main components of the learner module:

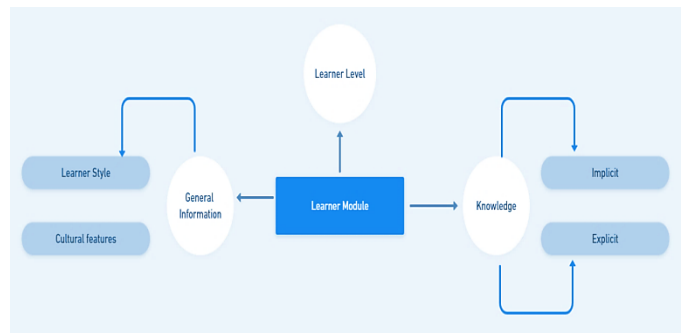


Fig. 1. Learner model component.

The analysis and study carried out in this section have enabled us to formulate various observations. The first one concerns the difficulties encountered by the learner module to extract implicit knowledge. For this, the authors have suggested a solution to convert tacit knowledge into explicit knowledge in the learner model of the intelligent tutoring system with the help of a competitive learning algorithm [12]. The second observation concerns the traces generated by the learners. The study conducted in this sense shows that the detection of learning styles is done automatically, without relying on explicit answers given to the questionnaire by the learners. Therefore, our contribution proposes an optimization of the activities suggested to the learner while keeping his preferred style and adding other parameters to achieve the required pedagogical objective.

III. METHODOLOGY

To reach our goal, we will apply evolutionary algorithms; and more precisely the genetic algorithm. Many researchers such as Goldberg, Davis and Michalewicz [13] have developed genetic algorithms (GA). GAs is certainly the best popular example of evolutionary algorithms [14][15][16][17]. The genetic algorithm is defined by a population cycle and involves three main factors: fitness, crossover, and mutation [3]. A population cycle represents the transition from one population to the next generation.

A. Level of Preference

In this subsection, we want to detect the learner's level of preference in each learning style. To do so, we rely on KOLB's model, which classifies the learner into four learning styles: Theorist, Activist, Reflective and Pragmatic. Each learner has a level of preference in a specific learning style. To do this, we prepare a list of activities $A_i = \{i=1...12\}$ each learning style (LS) contains activities ranging from 1 to 4 (input) and, the output indicates the three levels of preferences which is the total of activities chosen by the learner multiplied by the level of preference. The figure below shows an example of the preference level corresponding to the two learning styles (Activist and Reflectors).

Concerning the Pragmatist learning style, it contains two activities; the corresponding preferences are described in the Fig. 2 to 4.

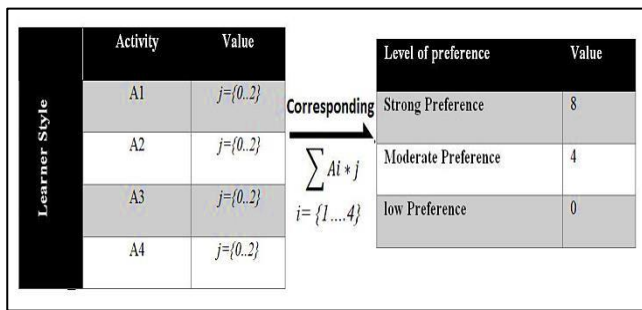


Fig. 2. Level of preference of Activist and Reflector’s learner style.

Concerning the Pragmatist learning style; it contains two activities; the corresponding preference is described in the figure below:

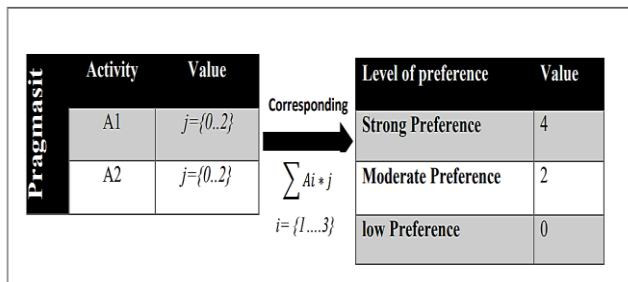


Fig. 3. Level of preference of the Pragmatist learner style.

The last learning style is theoretical; it contains three activities and three preferences levels.

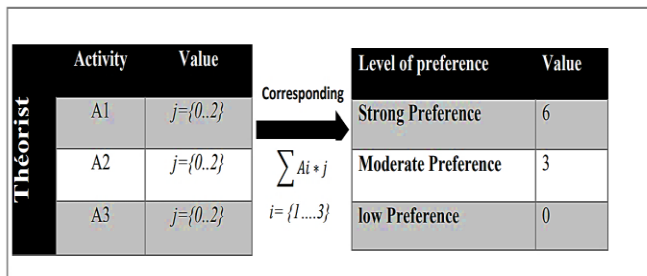


Fig. 4. Level of preference of the theorist learner style.

B. Description of the Genes

The proposed genetic algorithm detects the combination of actions that the student usually performs during a learning sequence using different learning activities. The production of the best-fit chromosome presents the combination of actions preferred by the student.

The first three genes of the chromosome represent the actions that correspond to the Theorist style defined by KOLB. The next four genes represent the actions corresponding to an activist style; the remaining four genes correspond to a reflector style, and the last two genes correspond to a pragmatist’s type. With the values of these genes, it is possible to establish the mechanisms of each learning activity that characterize a learner and, consequently, the optimization of these activities and the deduction of the learning style corresponding to the analysed chromosome.

The values of the genes g1, g2 and g3, carry information about the learner’s participation in the three activities. If the sum of the genes values is zero, it means that the learner does not participate in any of the three activities. On the other hand, if the sum of the codes for these genes is 6, it means that the learner has made extensive use of these activities and can be considered extremely theorist. If the addition of the two genes results are an intermediate value; the learner is moderately passive, neutral or moderately active, depending on the difference between the value obtained and the extremes. The Table I describes the genes:

TABLE I. DESCRIPTION OF THE GENES

Gene Number	Activities	Gene value	Description
Gene 1	Exercise of Analyzing	0	Does not make it
		1	Makes 50% of the proposed exercise
		2	Accomplish the activity
Gene 2	Exercise based on the extraction of conceptual model	0	Does not make it
		1	Makes 50% of the proposed exercise
		2	Accomplish the activity
Gene 3	Supervise a course (pdf, World, PPT)	0	Only reads the introduction
		1	Reads 50% of the course
		2	Accomplish the course
Gene 4	Collaborative learning	0	Does not participate
		1	Participate, but does not accomplish the task
		2	Accomplish the task
Gene 5	Activities based on games	0	Does not observe
		1	Makes 50% of the proposed activity
		2	Accomplish the activity
Gene 6	Simulation activities	0	Does not do it
		1	Makes 50% of the activity
		2	Accomplish the activity
Gene 7	Activities based on observation	0	Does not observe
		1	Makes 50% of the proposed activity
		2	Accomplish the activity
Gene 8	Access to a video course	0	Watch less than 10% of the video
		1	Watch 50% of the video
		2	finish the video
Gene 9	Production reports	0	Does not make it
		1	Realize 50% of the reports
		2	Accomplish the reports
Gene 10	Brain storming Activity	0	Does not make it
		1	Makes 50% of the activity
		2	Accomplish the activity
Gene 11	Realize the projects	0	Does not make it
		1	Realize 50% of the project
		2	Realize the project
Gene 12	Tutorials	0	Does not do it
		1	Makes 50% of the activity
		2	Accomplish the activity

C. Initial Population

An initial population is a group of chromosomes that represent possible solutions to the problem considered, in this case, as possible combinations of actions that a learner can perform. The size of the initial population and the level of diversity determine the quality of its coverage of the space of possible solutions. Based on the number of genes in the proposed chromosome structure and the number of possible values that each gene can have, it is possible to consider the complete solution space to be of the order of 3^{12} .

For the empirical evaluation of the approach, we used initial populations of two different sizes: 3000 chromosomes and 6000 chromosomes. These values were chosen because they are small compared to the whole space of possible solutions (represent 0.5% and 1% respectively); nevertheless, they allow obtaining a good level of diversity by the random generation of the initial population. The weight of initial population genes has a value of zero because the chromosomes have not yet been evaluated in terms of any academic unit performed by the learner.

D. Fitness Function

The fitness function evaluates each chromosome in a population and gives a score or fitness value to each chromosome based on their ability to solve the problem. In this problem, the fitness value of a chromosome is determined based on how close the chromosome's action combination is to the learner's preferred action combination. To evaluate a particular chromosome in a given population, the fitness function is based on the actions performed by the learner while participating in a specific learning activity.

Then, the fitness value of a chromosome C is defined as the product of the gene weights and represents the level of preference in the chromosome, as shown in Equations 1, 2 and 3. Based on our experience as trainers and teachers in the computer field, we notice that the application-based learning style is more profitable than theory and other types of learning.

$$\text{Fitness}(C) = \sum_{j=1}^{12} \alpha * x \quad (1)$$

$$\text{With: } \alpha = \begin{cases} -1 & \text{if } (j \in [1,3]) \\ 2 & \text{if } (j \in [4,6]) \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

$$\text{And: } x = \begin{cases} 0 & \text{if the activity is not achieved} \\ 1 & \text{if the half of the activity is achieved} \\ 2 & \text{if the activity is completely achieved} \end{cases} \quad (3)$$

E. Selection

Once chromosome fitness is calculated for a population, a selection method allows the algorithm to randomly select pairs of chromosomes for reproduction. In this contribution, we used three types of selection: uniform selection, roulette wheel selection and turn selection. Fig. 5 shows the example of the selected chromosome. The authors will detail every form in the experimental result section.

F. Crossover

The crossover is the result obtained when two chromosomes share their particularities. It allows the genetic

mixing of the population and the application of the principle of heredity of Darwin's theory. In this paper, the authors applied a simple crossover which they subdivided the chromosome on $\frac{1}{2}$. Fig. 6 shows the example of the results of chromosomes.

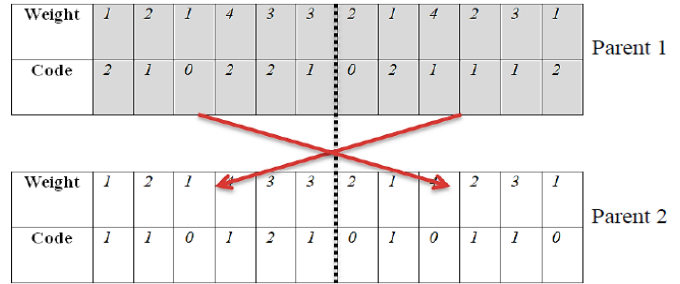


Fig. 5. Example of the selected chromosome (Parent1) & (Parent2).

Weight	2	1	4	2	3	1	1	2	1	4	3	3
Code	0	1	0	1	1	2	1	1	0	1	2	1

Descendants 1

Weight	1	2	1	4	3	3	2	1	4	2	3	1
Code	1	1	0	1	2	1	0	2	1	1	1	0

Descendants 2

Fig. 6. Example of the results chromosome (Descendants 1 and 2).

G. Mutation

The mutation is a genetic modification of the crossing chromosomes. The mutation is a method of introducing genetic diversity by changing the value or code of a randomly selected gene, as described in the Fig. 7. In this work, the authors decided to use the following function:

$$x_i = 2 - x_i \quad \text{With: } x_i = \{0,1,2\} \quad (4)$$

Weight	1	2	1	4	3	2	2	1	4	2	3	1
Code	1	1	0	1	2	1	0	1	0	1	1	0

Fig. 7. Example of a genetic mutation.

IV. RESULTS AND DISCUSSION

In order to evaluate the proposed approach, we simulated the actions performed at first by 10 learners generated arbitrary to build artificial data for the experiment. For this task, we considered that each learner has a particular learning style, represented by a set of preferred actions and that they behave according to this style. The actions performed by a learner correspond to their learning style. The resulting data is represented as a two-dimensional matrix, with each column representing the learner's identifier and an array of one particular action from the actions mentioned in the section above with twelve values.

TABLE II. DATA OF THE PREFERRED ACTIVITIES

Preferred Activities												
Learner ID	A 1	A 2	A 3	A 4	A 5	A 6	A 7	A 8	A 9	AI 0	AI 1	AI 2
1	0	1	2	2	1	2	1	2	0	2	1	2
2	2	2	0	1	0	1	1	0	0	0	0	0
3	0	2	0	1	0	1	1	2	2	2	2	0
4	2	0	2	0	2	2	0	2	2	1	2	0
5	0	1	1	0	1	2	1	1	0	0	0	0

The Table II illustrates the data from five learners selected randomly. The third row corresponds to the third learner, for example. This learner does not make the first activity. A2 indicates that he accomplishes the activity. The third activity marked that the learner read only the introduction, and he participates in the fourth activity, but he does not accomplish the task. A5 indicates that the learner does not observe the activity based on the game, but he makes 50% of the proposed activity in A6 and A7. The activities A8, A9, A10, and A11 indicate that the learner finishes the activities, but he does not do the last activity.

The Fig. 8 shows the evolution of the fitness function averaging 50 generations using the genetic algorithm uniform where the fitness value is 15. In the Mutation phase, if we change the value in the interval [4, 6], which has a coefficient (=2) that corresponds to an activist style, the fitness function decreases. On the other hand, if we change the value in the interval [1, 3], which has a coefficient (= -1) that corresponds to a theorist style, the fitness function progresses exponentially. This implies learning must tend towards the activist style.

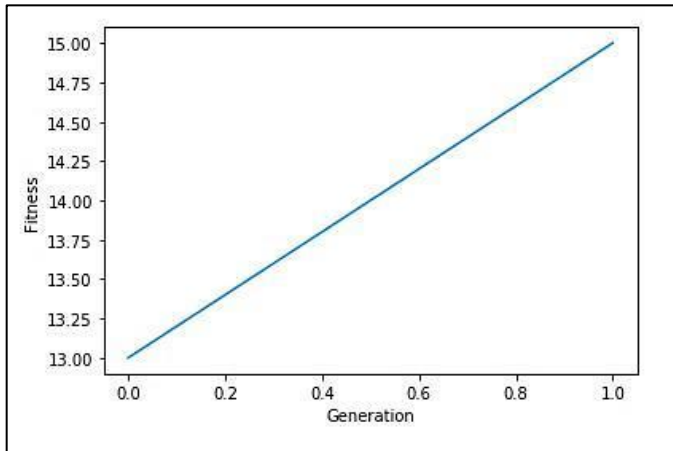


Fig. 8. Accuracy of genetic algorithm uniform.

Fig. 9 and 10 are the developed version of our proposed approach using the selection methods wheel roulette and tournament. We notice that the implementation of these methods improves the fitness function, which allows us to have good results. The fitness value is 21. The results show that the best selected population implies the improvement of the fitness function.

As long as the value of the fitness function is high, the learning tends towards the activist style. This entails the ignorance of theory-based activities and pragmatic activities,

which makes it possible to optimize the list of non-selected learning activities; and ultimately, improving learning process.

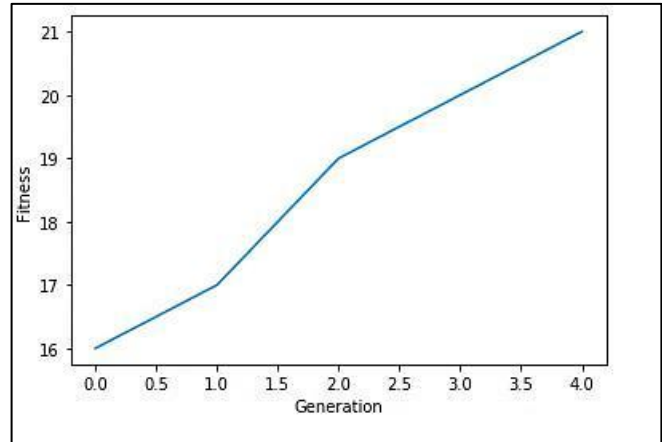


Fig. 9. Accuracy of genetic algorithm using wheel roulette selection.

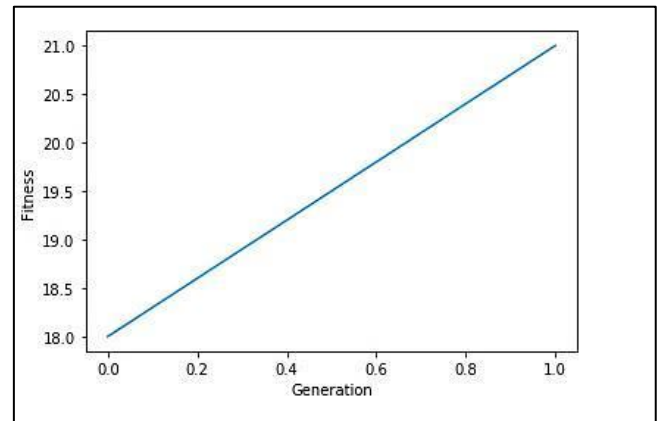


Fig. 10. Accuracy of genetic algorithm using tournament selection.

Table III shows that the 'Tournament' selection method performed better with the mean difference on the test set for the three data sets used compared to the 'Wheel roulette' selection method and the Uniform GA. Fig. 11 shows the variation of the selection methods in GA

TABLE III. COMPARATIVE STUDY BETWEEN THREE VERSIONS OF GA SELECTED METHODS

Comparisons	Mean Diff.	95.00% CI of diff.	Summary	Adjusted P Value
GA Uniform vs. GA-Wheel roulette	-4.000	-5.757 to -2.243	**	0.0017
GA Uniform vs. GA-Tournament selection	-5.500	-6.056 to -4.944	****	<0.0001
GA-Wheel roulette vs. GA-Tournament selection	-1.500	-2.812 to -0.1883	*	0.0308

From the previous researchers [18] and the results as in Fig. 12, the authors see that the tournament selection method (TSM) provides a high fitness value from the first generation compared to GA uniform (GA-U) and wheel roulette selection technique (WRSM). Nevertheless, both methods: TSM and WRSM, show good results at the end of the generation, fitness function =21 as shown in Fig. 12.

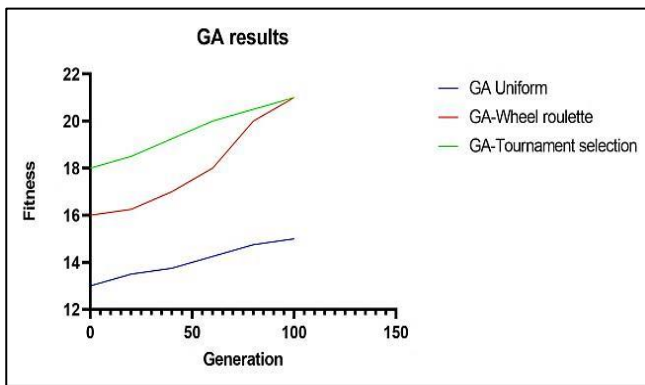


Fig. 11. Variation of the selection methods in GA.

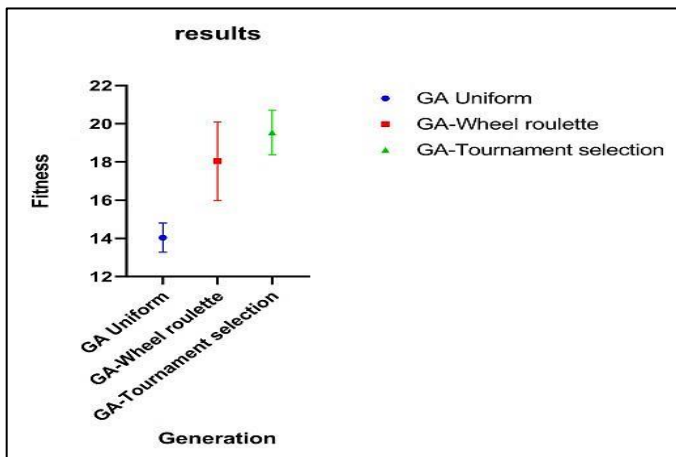


Fig. 12. Comparative result of three version of GA.

The recognition of learning style in an intelligent tutoring system increases the learning effectiveness. However, the adjustment of the content according to learner preferred style does not lead us to have achievable results. From the authors experience in the field of teaching, we recommend using the activist type and respecting the preferences of the learner.

According to the results obtained, we notice that the learners who use the practical activities acquire the expected objective rapidly. For this reason, we have developed and adapted the genetic algorithm by changing in the phase of mutation the activities based on the theory by low values x_i by respecting the following formula: $2-x_i$.

These results allow us to have an exponential function and converge to practice-based learning. These results urge the learner to perform well the case studies suggested by the intelligent tutoring system.

V. CONCLUSION

The e-learning system includes intelligent tools for the analysis, evaluation and assessment of the user's knowledge and skills, as well as for the monitoring and supervision of the e-learning process.

The study on data-driven approaches and more specifically on the traces generated by the learners shows that the detection of learning styles is done automatically, without relying on

explicit answers given to the questionnaire by the learners. Therefore, our contribution provides a path to be generated from the resource data (learner preference and activist style) towards the desired goal: acquiring the requested skills regarding the pedagogical objectives using genetic algorithms.

This contribution was developed to enhance the learning process of a programming language by presenting a new method for reducing the list of activities proposed to the learner. The authors are convinced that this contribution will lead to a good performance in building the learner model of the intelligent tutoring system. The implementation of these findings in the STS-programming solution is envisaged as future work.

REFERENCES

- [1] J. Ong, S. Ramachandran. "Intelligent tutoring systems: Using AI to improve training performance and ROI," Stotler Henke Associates, Inc. 2003.
- [2] Asselman, A., Khaldi, M. and Aammou, S. Evaluating the impact of prior required scaffolding items on the improvement of student performance prediction. *Educ Inf Technol* 25, 3227–3249. <https://doi.org/10.1007/s10639-019-10077-3>, 2020.
- [3] Fatima-Zohra Hibbi, Otman Abdoun and El Khatir Haimoudi. "Exploration of Analytical Mechanisms in the Feedback model", *Procedia Computer Science*, Volume 148, Pages 201-207, ISSN 1877-0509, 2019.
- [4] Hibbi, F.-Z., Abdoun, O., and Haimoudi, E. K. "Smart Tutoring System: A Predictive Personalized Feedback in a Pedagogical Sequence". *International Journal of Emerging Technologies in Learning (IJET)*, 16(20), pp. 263–268. <https://doi.org/10.3991/ijet.v16i20.24783>, 2021
- [5] Hmedna, B., El Mezouary, A. and Baz, O. "A predictive model for the identification of learning styles in MOOC environments". *Cluster Comput* 23, 1303–1328, <https://doi.org/10.1007/s10586-019-02992-4>, 2020.
- [6] Dunn, Rita, et al. "A Meta-Analytic Validation of the Dunn and Dunn Model of Learning-Style Preferences." *The Journal of Educational Research*, vol. 88, no. 6, Taylor & Francis, Ltd, pp. 353–62, <http://www.jstor.org/stable/27541998>, 1995
- [7] Felder, Richard. "Learning and Teaching Styles in Engineering Education", *Journal of Engineering Education*, vol. 78(7), 674–681, 1988.
- [8] Sabine Graf, Tzu-Chien Liu, Kinshuk, Nian-Shing Chen, Stephen J.H. Yang. "Learning styles and cognitive traits – Their relationship and its benefits in web-based educational systems", *Computers in Human Behavior*, Volume 25, Issue 6, Pages 1280-1289, ISSN 0747-5632, 2009.
- [9] K. Chrysafiadi and M. Virvou. "Student modelling approaches: A literature review for the last decade," *Expert Systems with Applications* 40 4715–4729. 2013.
- [10] E. Sierra, R. García-Martínez, Z. Cataldi, P. Britos, and A. Hossian, "Towards a methodology for the design of intelligent tutoring systems," *Research in Computing Science Journal*, vol. 20, pp. 181–189, 2006.
- [11] Fatima-Zohra Hibbi, Otman Abdoun and El Khatir Haimoudi, "Bayesian Network Modelling for Improved Knowledge Management of the Expert Model in the Intelligent Tutoring System" *International Journal of Advanced Computer Science and Applications(IJACSA)*, 13(6), 2022.
- [12] Hibbi, F., Abdoun, O., and El Khatir, H. "Extract Tacit Knowledge in the Learner Model of the Smart Tutoring System. *International Journal Of Emerging Technologie in Learning (IJET)*, 15(04), PP.23-240. Doi: <http://dx.doi.org/10.3991/ijet.v15i04.11781>", 2020
- [13] Goldberg, David Edward. "Genetic algorithms in search, optimization, and machine learning", Addison-Wesley, 412 p, 1989.
- [14] Bck T. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York, 1996.

- [15] J. Dro, A. Ptrowski, P. Siarry, and E. Taillard. "Mta-heuristiques pour l'optimisation difficile". Eyrolles, 2003.
- [16] O. Abdoun and J. Abouchabaka. "A Comparative Study of Adaptive Crossover Operators for Genetic Algorithms to Resolve the Travelling Salesman Problem". *International Journal of Computer Applications*, Vol.31, N.11, 2012.
- [17] Yassine Moumen, Otman Abdoun, and Ali Daanoun. "Parallel approach for genetic algorithm to solve the Asymmetric Traveling Salesman Problems". In *Proceedings of the 2nd International Conference on Computing and Wireless Communication Systems*. ACM, 2017.
- [18] Jinghui Zhong , Xiaomin Hu , Min Gu and Jun Zhang, (2005). "Comparison of Performance between Different Selection Strategies on Simple Genetic Algorithms", *Proceedings of the International Conference on CIMCA-IAWTIC'05*.

Algorithm for Skeleton Action Recognition by Integrating Attention Mechanism and Convolutional Neural Networks

Jianhua Liu*

College of Physical Education, Weifang University, Weifang, 261061, China

Abstract—An action recognition model based on 3D skeleton data may experience a decrease in recognition accuracy when facing complex backgrounds, and it is easy to overlook the local connection between dynamic gradient information and dynamic actions, resulting in a decrease in the fault tolerance of the constructed model. To achieve accurate and fast capture of human skeletal movements, a directed graph convolutional network recognition model that integrates attention mechanism and convolutional neural network is proposed. By combining spacetime converter and central differential graph convolution, a corresponding central differential converter graph convolutional network model is constructed to obtain dynamic gradient information in actions and calculate local connections between dynamic actions. The research outcomes express that the cross-target benchmark recognition rate of the directed graph convolutional network recognition model is 92.3%, and the cross-view benchmark recognition rate is 97.3%. The accuracy of Top-1 is 37.6%, and the accuracy of Top-5 is 60.5%. The cross-target recognition rate of the central differential converter graph convolutional network model is 92.9%, and the cross-view benchmark recognition rate is 97.5%. Undercross-target and cross-view benchmarks, the average recognition accuracy for similar actions is 81.3% and 88.9%, respectively. The accuracy of the entire action recognition model in single-person multi-person action recognition experiments is 95.0%. The outcomes denote that the model constructed by the research institute has higher recognition rate and more stable performance compared to existing neural network recognition models, and has certain research value.

Keywords—Attention mechanism; convolutional neural network; action recognition; central differential network; spacetime converter; directed graph convolution

I. INTRODUCTION

Human Action Recognition (HAR) is a research field that has received much attention from scholars both domestically and internationally. HAR plays an important role in public places, healthcare, and safety [1-2]. Early research on HAR was generally based on RGB videos. With the development of technology, algorithms and sensors for estimating human skeletal posture have emerged [3]. Researchers use depth cameras to extract depth information from human skeletal information at different scales, lighting, and backgrounds to construct HAR data in a 3D coordinate system, and construct corresponding algorithm models based on deep learning (DL) [4]. In the research of HAR using 3D skeleton data as the experimental object, HAR and feature capture often use 3D

technology to obtain relevant action data information [5]. However, there are still many problems in the rapid acquisition of human actions in related operations. The changes and complexity of the background and environment in the video can result in different skeletal joint information for the same action, and the mutual occlusion between people and background can also reduce the extraction of feature actions. The already built HAR model may also have problems, such as identifying only simple action types, making it difficult to identify more complex behaviors. The above issues will result in the already constructed HAR model being only usable under specific conditions, and the experimentally constructed model cannot be generalized to practical applications. To improve the recognition accuracy of action recognition (AR) models and eliminate interference caused by factors such as environment, background, and occlusion, a directed graph convolutional network (DGCN) recognition model for enhancing attention was proposed. By recognizing the spacetime information extracted from feature extraction, a central differential converter graph convolutional network was constructed to achieve real-time recognition and capture of human actions. The Section II of the study mainly discusses the relevant research on human action models in recent years. The Section III mainly constructs an enhanced information acquisition and enhanced spatiotemporal information conversion graph convolution model. The Section IV tests the performance of the constructed model using different datasets and compares the effectiveness of different HAR models on the same dataset. Section V discusses the results. The Section VI summarizes the research results and draws research conclusions.

II. RELATED WORKS

Many scholars have made achievements in the construction of HAR models. Gu et al. constructed a new HAR model using the improved sparse classification model and deep convolutional neural network (CNN), and applied the model to the benchmark dataset. The performance of the model was verified through setting experiments [6]. Huang et al. utilized neural networks for pseudo-image processing of skeletal data. A novel CNN with an adaptive inference framework was constructed by utilizing the dependent joints between skeletal joints [7]. To improve the detection accuracy of human bone models, Li proposed a multi-branch multi-level cascaded CNN structure model to predict the information of occluded parts [8]. Peng et al. proposed a three stream model using two different types of deep CNNs to improve the generalization of HAR models, and verified the effectiveness of the model by

experiments [9]. Yang put forward an algorithm model based on random projection combined with multi-channel 3D CNN to improve the accuracy and recognition speed of HAR models. The effectiveness of the algorithm model was demonstrated through experiments [10]. To raise the accuracy of human action model recognition in video segmentation and its computational efficiency in large-scale datasets, Zhao et al. proposed a multi-dimensional data model for video image AR and non-action based on DL framework, and verified the effectiveness of the model through experiments [11]. Liu and Che used attention spacetime convolutional graph networks to learn the importance of different actions in sports videos and analyze different actions in sports videos [12]. To solve the problems of background clutter, scene diversity, viewpoint change, occlusion, etc. in the HAR, He et al. put forward a closely connected bidirectional long and short-term memory (LSTM) network DL model to capture the temporal and spatial patterns of human action in videos. The research findings also indicated that the effectiveness of the proposed model was good [13]. Wenbo et al. put forward a protocol recognition method based on CNNs to improve the accuracy of feature acquisition, and it was proved that the proposed algorithm had high accuracy and fast convergence speed [14]. Ma et al. proposed a novel deep convolutional generative adversarial network to recognize human action posture, and verified through experiments that its model had significant advantages over existing models [15]. Chen et al. put forward a multi-radar collaborative HAR model based on transfer and integrated learning to solve the view limitation in AR. Through experiments, the proposed model had higher recognition accuracy compared to the single-view radar fusion model [16]. To promote the accuracy of action combination training AR model, Jiang and Tsai proposed a model based on sequential minimal optimization model and artificial intelligence. And in subsequent experiments, it has been proven that this model could improve the recognition rate of actions and meet the recognition needs of online actions [17]. To establish a better HAR model, Chen et al. proposed a model that integrated LSTM weight convolution neural networks. It was validated that the model had an authentication accuracy of 98.0% [18].

In summary, research on constructing HAR models using CNNs has become mature, but there is still room for improvement in accuracy. Based on this, a CNN human recognition model integrating attention mechanism (AM) is proposed, which combines central difference graph convolution network (CDGN) and spacetime converter (SC) to achieve accurate and fast human actions recognition.

III. THE CONSTRUCTION OF SKELETON AR MODEL BASED ON AM AND CNN

This study mainly uses graph convolutional neural networks (GCNN) as the skeleton and integrates AMs to design a DGCN model to enhance the recognition and capture of action information. To prevent the omission of dynamic gradient information in actions and calculate local connections between dynamic actions, CDGN and SC are introduced to construct an enhanced graph convolution model for spacetime information conversion.

A. The Construction of ADGCN Skeleton AR Model Integrating AM and Neural Network

HAR, as an emerging research project in computer vision, has great impact on many fields. Medical assistance, human-computer interaction, intelligent driving, and intelligent security can be achieved through the recognition of HAR. The method of skeleton AR is generally completed using 3D skeleton data, using skeleton joint points to form corresponding groups, and then connecting the corresponding groups to complete the representation of the entire body structure. At present, the commonly used methods for HAR include two types: extraction based on traditional manual features and extraction based on DL methods. DL methods mainly include methods based on recurrent neural network (RNN), CNN, fusion based on RNN and CNN, and GCNN. GCNN takes skeleton data modeling as a blueprint, joints as fixed points, and draws skeleton edge maps. Compared to traditional manual feature extraction and CNN and RNN methods, it has a more intuitive description of skeleton data. Therefore, the study selects the GCN method as the basis to construct a model that is more convenient for skeleton AR. The flow chart of GCN object diagram convolution is shown in Fig. 1.

From Fig. 1, the GCN method involves the temporal structure and continuity of human skeleton actions recognition. By decomposing a certain action information into corresponding data, and then placing the decomposed data into different channels according to dimensional relationships for processing and operation, the final action analysis result is obtained. Due to the existing skeleton models constructed based on GCN, their skeleton AR accuracy may be reduced in complex backgrounds and dynamic environments. And the skeleton model constructed by GCN is often an undirected graph, which can only determine whether there is a connection relationship between adjacent vertices and edges in the graph, and cannot capture more feature information.

To address the above issues, an AM is introduced to improve the GCN skeleton model. A multi-stream framework is used to capture the positions of joints and bones, identify their action trajectories and features, establish the dependency relationship between vertices and edges in the skeleton graph, introduce an AM to focus on feature channels, and construct an attention enhanced direct graph convolutional network (ADGCN) skeleton AR model. The workflow diagram of the ADGCN model is displayed in Fig. 2.

From Fig. 2, the ADGCN skeleton AR model is divided into five parts: multi-stream framework input, DGCN processing, attention enhancement network, data fusion, and output data. The data input by the multi-stream framework corresponds to the action information of joints and bones. The related motion information is judged by the adjacency of DGCN and captured by the key frame action of the attention enhancement network. The data processed by the first four information flow are weighted and fused to complete the prediction of related actions. The update and aggregate functions are used to determine the connection of vertices and edges in the ADGCN model to represent the connected vertices between two vertices and the relationship between vertices and

edges after edge update. The update function of vertices is shown in equation (1).

$$v_i' = u^v([v_i, e_j, e_j]) \quad (1)$$

In equation (1), v_i' is the updated vertex, v_i is the vertex; $[\cdot]$ means the connection operation; u indicates the update function; e_j, e_j denote the position of the updated vertex v_i' and its adjacent edge e . Similarly, it obtains the function equation (2) between the updated edge e' and adjacent vertices.

$$e_j' = u^e([e_j, v_j^s, v_j^t]) \quad (2)$$

In equation (2), v_j^s, v_j^t stand for the updated source vertex and target vertex, respectively. Let the joint coordinate at time t be $v_{i,t} = (x_{i,t}, y_{i,t}, z_{i,t})$, and the calculation function for the same node in two consecutive frames is shown in equation (3).

$$m_{i,t+1} = v_{i,t+1} - v_{i,t} \quad (3)$$

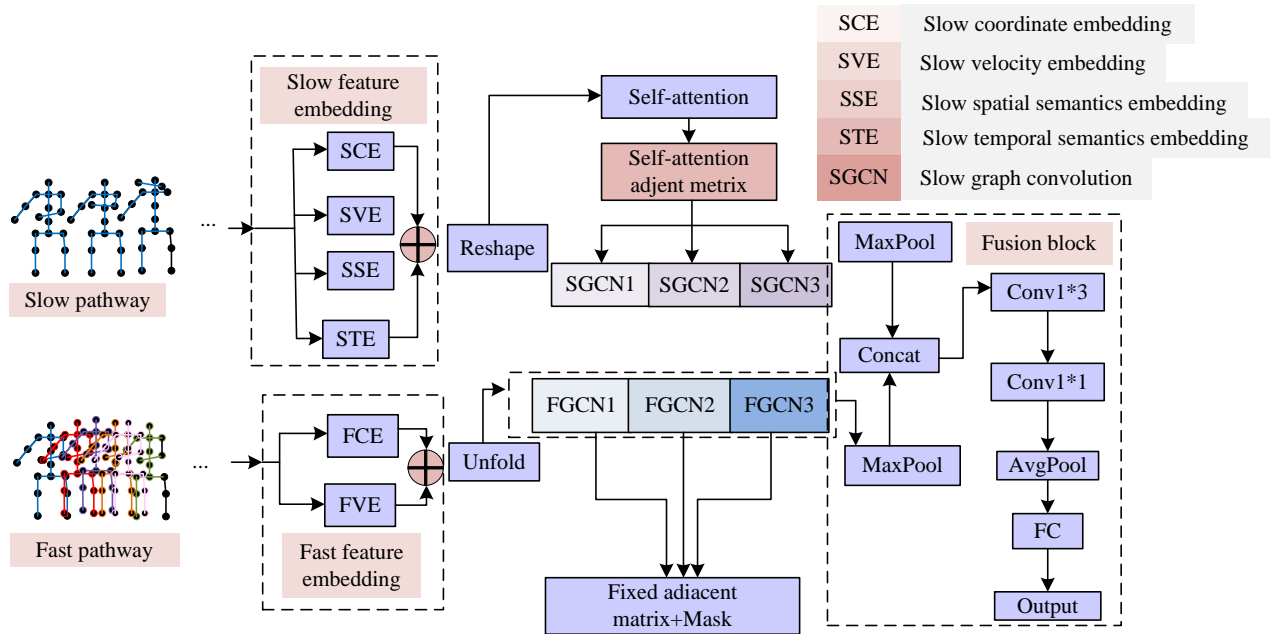


Fig. 1. Flow chart of GCN object diagram convolution.

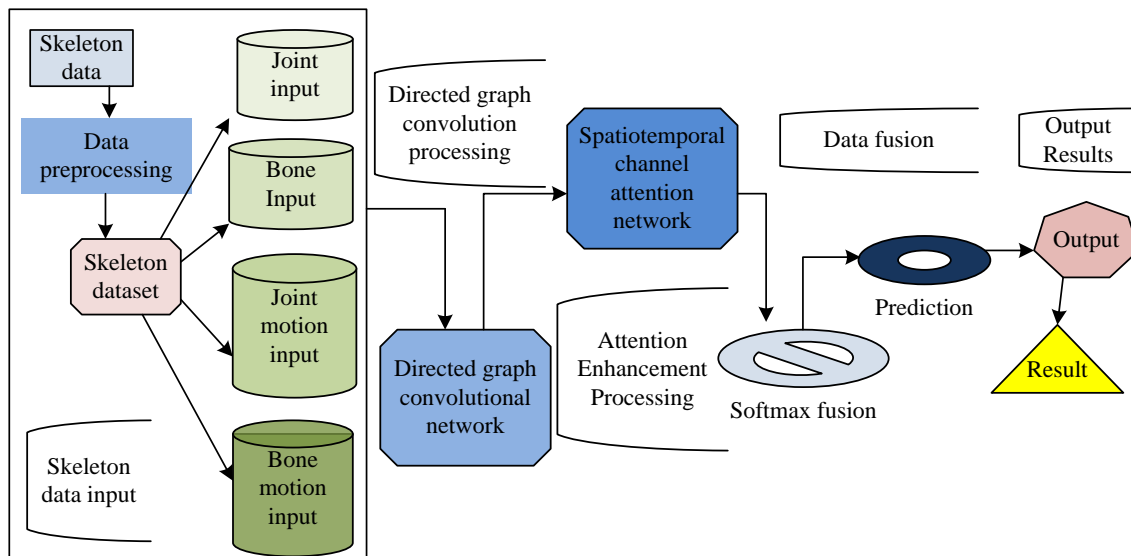


Fig. 2. ADGCN model workflow diagram.

Designing a spacetime channel attention network (CAN) is to strengthen the information connection between key joints and bones, while taking care not to weaken the information intensity of other joints. The spacetime CAN constructed by the research institute includes three types: spatial attention network (SAN), temporal attention network (TAN), and CAN. Among them, SAN mainly allocates different degrees of attention to joints and the bones connected to them. The attention of vertices and edges in the graph is M_s , and the set of M_s is expressed as $\{M_{s,v}, M_{s,e}\}$. The corresponding calculation method is shown in equation (4).

$$M_s = \delta(g_s(\text{AvgPool}(fin))) \quad (4)$$

In equation (4), δ means the sigmoid activation function; g_s indicates a one-dimensional convolution operation; fin refers to the input data information of vertices and edges in the directed graph. As a dynamic time selection mechanism, TAN mainly determines when attention begins and sets the attention time of vertices and edges in a directed graph to M_t . The relevant function expressions are shown in equation (5).

$$M_t = \delta(g_t(\text{AvgPool}(fin))) \quad (5)$$

In equation (5), $M_t \in R^{1 \times T \times 1}$, the explanation of parameter meanings refers to equation (4). After allocating time attention and attention to joints and edges, it considers enhancing the model's feature description of input samples, and thus introduces CAN. The relevant calculation is shown in equation (6).

$$M_c = \delta(W_2(\tanh(W_1(\text{AvgPool}(fin)))) \quad (6)$$

In equation (6), W_1 and W_2 are the weights of two full connection layers, and the functions used by TAN network are \tanh and Sigmoid activation functions. SAN, TAN, CAN three kinds of attention enhancing networks are serially arranged to enhance the recognition ability of spacetime CAN and highlight its attention enhancement effect.

B. Constructing a Graph Convolutional Model for Enhanced Spatiotemporal Information Transformation Based on CDTN

After the construction of the enhanced information acquisition model, its ability to obtain spacetime channel information has been raised to a certain degree, but it ignores the dynamic gradient information in the actions and the local connections between the corresponding dynamic actions. Introducing CDGN can further improve the model constructed by the research institute, by obtaining corresponding dynamic gradient information and using a converter to obtain fixed point connections between nodes. CDGN, as an image processing algorithm, is mainly used for image edge detection and feature extraction. It uses the difference between the central and adjacent pixels to calculate the value of new pixels, reducing noise interference in the image and enhancing its features. The CDGN algorithm is shown in equation (7).

$$\begin{cases} g_x(x, y) = f(x+1, y) - f(x-1, y) \\ g_y(x, y) = f(x, y+1) - f(x, y-1) \end{cases} \quad (7)$$

In equation (7), $g_x(x, y)$ and $g_y(x, y)$ express different gradient values in the horizontal and vertical directions, respectively, and $f(x, y)$ means the pixel values in the original image. The edge values and features of the image are calculated by calculating the gradient value of each pixel point. CDGN includes two parts: sampling and aggregation, and its feature aggregation is shown in Fig. 3.

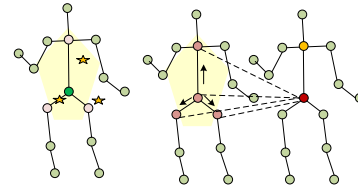


Fig. 3. CDGN feature aggregation principle.

From Fig. 3, the sampling of CDGN takes into account both feature differences and gradient features. After taking into account the differences between certain node and central node and the gradient features of the corresponding central node, the connection is made according to the corresponding gradient direction. From this, the center gradient of the sampling vertices is aggregated and represented by equation (8).

$$O(v_i) = \sum_{v_j \in Ri} \frac{1}{Z_{ij}} w(l_i(v_j)) * (I(v_j) - I(v_i)) \quad (8)$$

In equation (8), I indicates the input feature; O means the output feature; w indicates the weight function; Ri expresses the first order adjacency joint distance of vertex v_i ; l_i is the partition function. The converter network can coordinate the global self-attention, self-attention operation and convolution operation. Combining the CNN with the converter can allow the model to mine more relevant information from the captured features. The SC used in the research institute consists of three parts: joint embedding, SC attention module, and time converter attention module. The network structure of the SC is shown in Fig. 4.

From Fig. 4, different modules implement different functions. Among them, the joint marker embedding module includes a multi-head self-attention (MHSA) layer and a feedback neural network (FNN) layer to extract spacetime features. The SC attention module is applied to the acquisition of labeled space, while the time converter attention module corresponds to the calculation of spatial and temporal dependencies. Combining CDGN and SC network models, a central difference transformer graph network (CDTG) is developed to capture action gradient information and calculate local dependencies between nodes. The CDTG network structure is shown in Fig. 5.

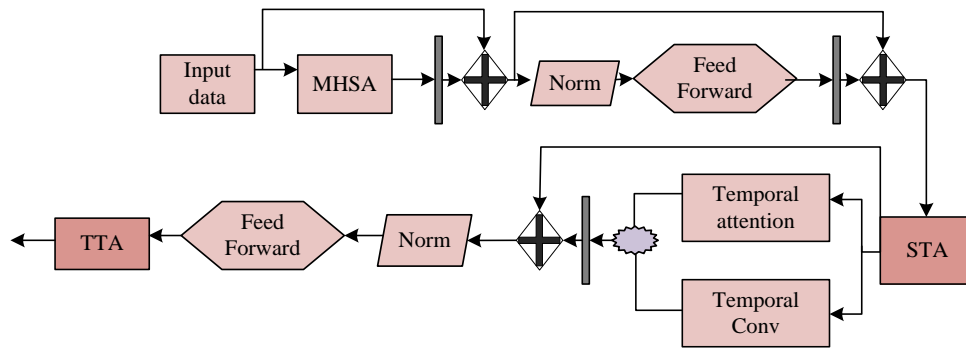


Fig. 4. Diagram of SCNetwork.

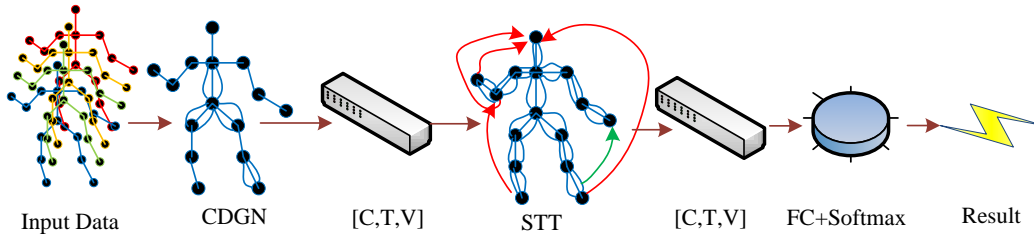


Fig. 5. CDTG Network architecture.

From Fig. 5, CDTG is divided into four parts: skeleton data input, CDGN, SC attention network, and fully connected layer recognition information. The last one will use the Softmax function, and the relevant expressions are shown in equation (9).

$$P(y = j) = \frac{e^{x^T W_j}}{\sum_{k=1}^K e^{x^T W_k}} \quad (9)$$

In equation (9), $P(y = j)$ expresses the probability that the sample x belongs to the j classification; k denotes the result of linear function, and is the weight value. The expression of cross entropy function is shown in equation (10).

$$H(p, q) = \sum_x p(x) \log q(x) \quad (10)$$

In equation (10), p means the true labeled distribution, and q refers to the pre labeled distribution of the trained model. It inputs the skeleton action information into CDGN, updates the spatial gradient information of the next layer's vertices based on the position of the central node and its adjacent nodes. The dependency relationship of the global spacetime nodes is calculated based on the attention network of the SC, and finally the data are processed through the fully connected layer recognition information processing and weighted average to complete the prediction of the entire action behavior.

IV. ANALYSIS OF AR RESULTS FOR ADGN AND CDTG MODELS

The experiment used NTU-RGB+D, MSR-Action 3D, and SBU datasets collected by the Kinect v2 3D tactile camera as the research dataset. Experiments were organized to evidence the effectiveness of the proposed ADGCN algorithm model. The NTU-RGB+D dataset contained 56880 skeleton action video sequences, with each skeleton action video covering 25 nodes and each node providing corresponding 3D coordinate positions. MSR-Action 3D contained 20 types of actions and 567 data sequences. The SBU dataset contained 8 types of actions, 284 videos, and 15 skeleton joint points. The environmental parameters for the skeleton AR experiment are shown in Table I.

TABLE I. EXPERIMENTAL PARAMETER SETTINGS FOR SKELETON AR

Experimental setup	Experimental parameters
Experimental system	Linux Ubuntu18.04
GPU	GeForce RTX2080Ti
Computing Platform	CUDA10.0

The batch size of the model is set to 64 and the initial weight attenuation value to 0.0005. The model was trained in the NTU-RGB+D dataset. The accuracy of the trained model was tested on the NTU-RGB+D dataset for crosstarget and view benchmarks. And it compared the accuracy of manual feature extraction algorithms Lie Group, RNN-based feature extraction methods (GCA-LSTM and STA-LSTM), CNN-based methods (3SCNN and TCN), and GCN-based methods (ST-GCN, 2sAGCN, DGCN) on CV and CS. The research outcomes are expressed in Fig. 6.

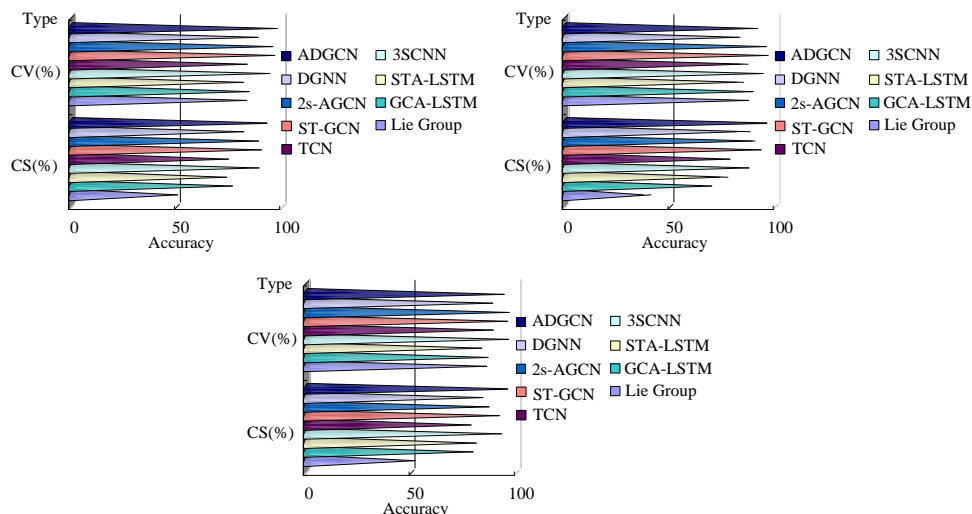
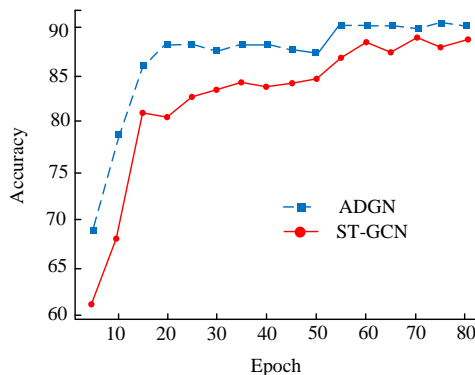


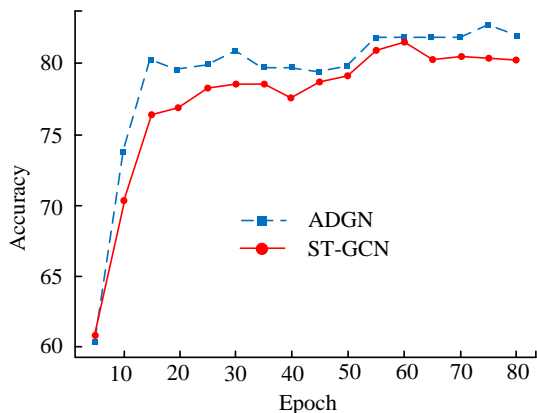
Fig. 6. Performance comparison between algorithms on the NTU RGB+D dataset.

From Fig. 6, in the three datasets of NTU RGB+D, MSR-Action 3D, and SBU, the artificial feature extraction method Lie Group had the lowest CS recognition accuracy, with 50.1%, 47.5%, and 50.5%, respectively, while other DL-based methods had an accuracy of over 73% in CS benchmark recognition. The CS accuracy based on the GSCN network structure was over 80%, with the highest being ST-GCN89.90%. The ADGCN model proposed by the research institute scored the highest among the four network models, with CS benchmark recognition accuracy of 92.33%, 93.21%, and 94.12% in the three datasets, respectively. In cross-view benchmark recognition, the recognition accuracy of the four algorithms was higher than that of cross-target benchmark recognition. The model based on the GSCN network structure still had higher CV recognition accuracy than the other three network models, while the ADGCN model proposed by the research institute had the highest CV recognition accuracy in the past, with 97.3%, 97.8%, and 96.7% in the three datasets, respectively. To further validate the effectiveness of the model, the ST-GCN model with the second highest score was selected to compare the results of the proposed ADGN model CS and CV benchmark recognition accuracy changes with the training. The laboratory findings are shown in Fig. 7.



(b)Comparison of contrast rates between two models under CV

Fig. 7. Changes in the accuracy of CS and CV benchmark recognition for two algorithms during training.



(a)Comparison of contrast rates between two models under CS

From Fig. 7, as the number of iterations continued to increase, the accuracy of the two bone AR models would show an upward trend. Under the CS benchmark, after the number of iterations reached 15, the recognition accuracy of the model tended to stabilize. However, the recognition accuracy of the ADGN model has always been higher than that of the ST-GCN model, with a maximum difference of 0.5% and a maximum difference of 3.42%. Under the CV benchmark, the recognition accuracy of both models has increased compared to the CS benchmark, with an increase of about 5%. After 15 iterations, the recognition accuracy of the model tended to a relatively stable accuracy range, with the maximum recognition accuracy of ADCN being 90%, ST-GCN being 88.5%, and the ADCN recognition accuracy curve consistently above ST-GCN. The experimental results indicated that the ADCN model constructed by the research institute had a certain degree of stability and could ensure good recognition accuracy. Experiments were designed to verify the recognition accuracy of the ADGCN model for different levels of actions on the Kinetics Skelton dataset, and it compared the recognition accuracy of Deep LSTM, TCN, DGCN, and SAN algorithms for the same action. The research outcomes are displayed in Fig. 8.

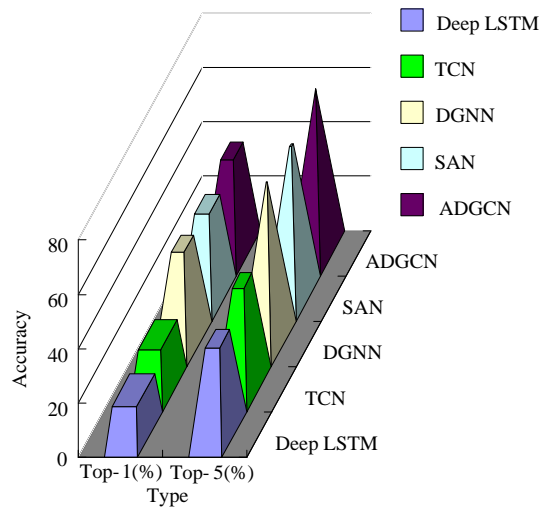


Fig. 8. Comparison of recognition accuracy of the kinetics skeleton dataset.

Due to the fact that the action videos in the Kinetics Skelton dataset are sourced from YouTube, with a larger variety and complexity compared to the previously used NTU RGB+D dataset, the recognition accuracy of the five validated models is lower. From Fig. 8, in the type of Top-1, the recognition accuracy of the SAN, DGNN, and ADGCN models was around 35%, belonging to the three branches with good performance among the five algorithms. The ADGCN model's Top-1 AR accuracy was higher than DGNN with a 0.7% advantage, ranking first. Compared to Top-1, in Top-5 indicator recognition, the recognition accuracy of the five research algorithms has improved, with an increase of 18.6% to 23.0%. Among them, DGNN had the largest increase, the Top-5 accuracy of the model was 56.5%, and ADGCN had an increase of 22.9%. However, the recognition accuracy of ADGCN was the highest, with 60.5%. It designed experiments to evidence the effectiveness of the CDTG model proposed by the research institute. In the NTU-RGB+D dataset, it compared the recognition accuracy obtained by CNN-based methods (HCN, TCN, GCNN, Clips+ CNN+MTLN), RNN-based methods (ST-LSTM, LSTM-CNN), and GCN-based methods (ST-GCN STGR-GCN GR-GCN GCST Dynamic GCN). The research findings are expressed in Fig. 9.

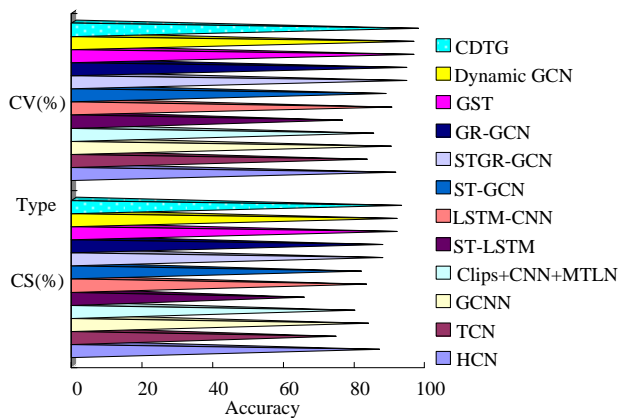
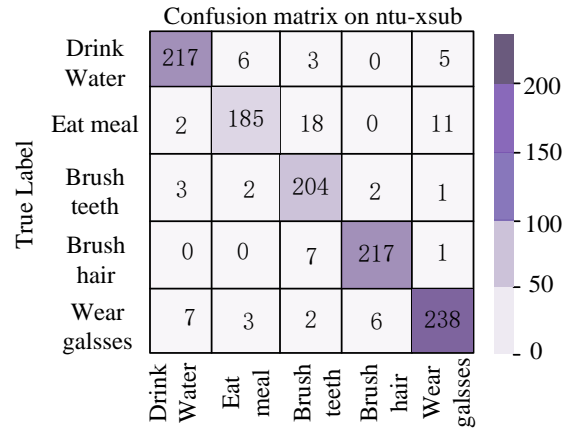
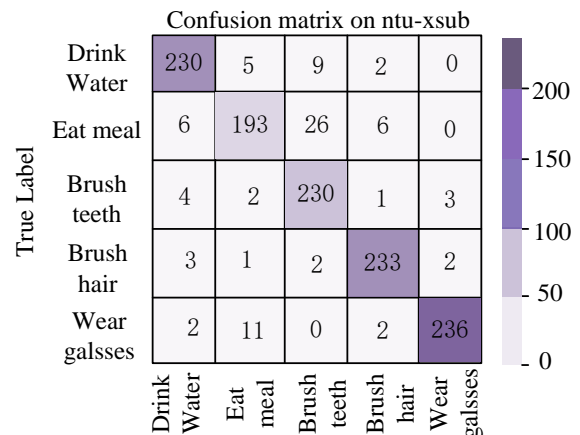


Fig. 9. Comparison of recognition accuracy between different algorithms.

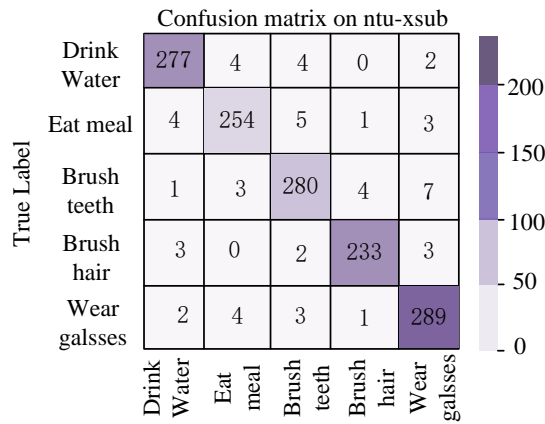
From Fig. 9, the recognition accuracy of the model based on GCN network structure on CS and CV benchmarks was higher than that of the model built on CNN and RNN. Among the models constructed using GCN networks, GCST, Dynamic GCN, and the CDTG model constructed by the research institute had relatively good recognition accuracy. Among them, the CDTG model had the highest recognition accuracy, with recognition accuracy rates of 92.87% and 97.52% in CS and CV, respectively. Compared with other types of AR models, its accuracy improvement could reach a maximum of 27.67% (CS benchmark) and 21.42% (CV benchmark). Compared to the GCST and Dynamic GCN models with higher AR accuracy in recent years, the recognition accuracy of CDTG in CS benchmark had increased by 1.36% and 1.37%, respectively. The recognition accuracy of CV benchmark has been improved by 1.32%. This indicated that the CDTG model proposed by the research institute had good recognition performance in skeleton AR. To further evidence the recognition performance of the CDTG model, five categories with similar action content were selected for comparison in the NTU-RGB+D dataset. The selection of similar action content was divided into five categories: "drinking water", "eating", "brushing teeth", "washing hair", and "wearing glasses". The experimental results were set up to verify the recognition results of CDTG and ST-GCN for the above five action categories. The experimental results are shown in Fig. 10.



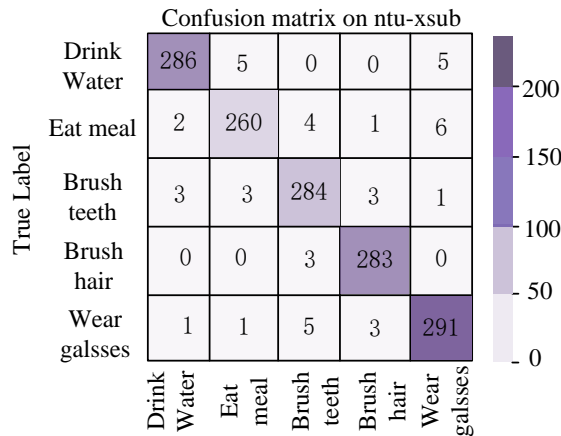
(a) Prediction results of ST-GCN under CS



(b) Prediction results of CDTG under CS



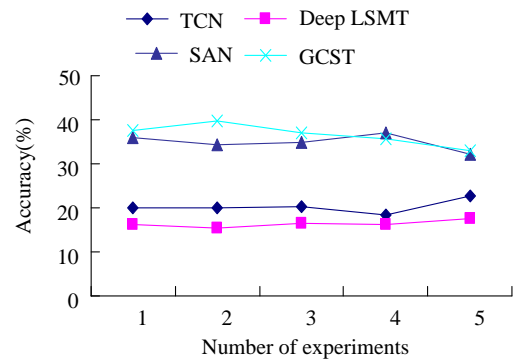
(c) Prediction results of ST-GCN under CV



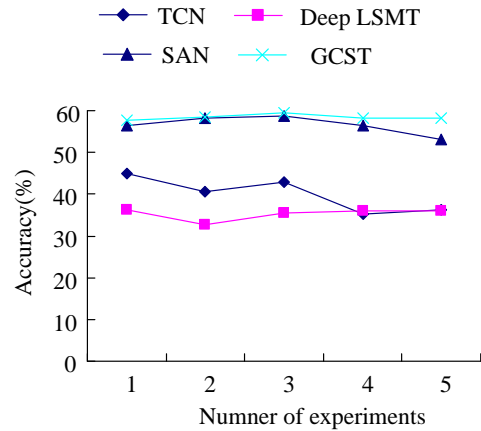
(d) Prediction results of CDTG under CV

Fig. 10. Comparison of recognition results of similar actions between ST-GCN and CDTG models.

From Fig. 10, the recognition performance of the two models under the CV benchmark was better than that under the CS benchmark. There were 276 CS benchmark samples to be tested, among which the recognition numbers for the five actions of "drinking water", "eating", "brushing teeth", "washing hair", and "wearing glasses" in ST-GCN were 217, 185, 204, 217, and 238, respectively, with recognition accuracy rates of 78.6%, 67.0%, 73.9%, 78.6%, and 86.2%. The average recognition rate was 76.9%. The recognition rates of the CDTG model for the above five actions corresponded to 83.3%, 69.9%, 83.3%, 84.4%, and 85.5%, with an average recognition rate of 81.3%. The number of samples to be tested under the CV benchmark was 316, indicating that the recognition accuracy of the CDTG model was better than that of ST-GCN. The recognition rates for "drinking water", "eating", "brushing teeth", "washing hair", and "wearing glasses" actions were 90.5%, 82.3%, 89.9%, 89.6%, and 92.1%, respectively. The average rate of AR was 88.9%. The average recognition rate of ST-GCN was 86.3%. It designed experiments to verify the recognition accuracy of CDTG on Top-1 and Top-5 in the Kinetics Skelton dataset, and compared the recognition accuracy of TCN, Deep LSTM, and SAN GCST models. The experimental results are shown in Fig. 11.



(a) Comparison of Model Performance Regarding Top-1 Indicators



(b) Comparison of Model Performance Regarding Top-5 Indicators

Fig. 11. Comparison of accuracy of five algorithms on the kinetics skeleton dataset.

From Fig. 11, the CDTG model had better recognition accuracy in Top-1 and Top-5 compared to other models, with the highest accuracy among the comparison models, at 37.58% and 59.61%, respectively. Compared to other types of AR models, its Top-1 had a maximum improvement accuracy of 21.18%. The minimum improvement accuracy was 0.88%. The highest improvement in recognition accuracy of Top-5 was 24.31%, and the lowest improvement accuracy was 1.25%. Four experimental testers (two males and two females) were identified and predicted for 3D movements using ADGN and CDTG models. The research outcomes are shown in Table II.

TABLE II. RESULTS OF ADGN AND CDTG MODELS ON PHYSICAL EXPERIMENTS

Action label	Drinking	Smoking	Walking	Jumping
Male tester 1	True	True	True	True
Male tester 2	True	True	True	True
Female tester 3	True	False	True	True
Female tester 4	True	True	True	True
Action label	Shaking hands		Hugging	
Tester 1+2	True		True	
Tester 3+4	True		True	

From Table II, the combination model constructed by the research institute can accurately recognize single and multi-person movements, with a recognition accuracy of 100% for "drinking water", "walking", and "jumping" movements. However, in the recognition of "smoking" movements, female test subject 3 made an error in identifying the movements, which might be due to the insufficient fine-grained model constructed by the research institute. On the other hand, it might also be that female testers were not familiar with the 'smoking' action, which led to model recognition errors. The experimental results indicated that the action model constructed by the research institute had good recognition performance.

TABLE III. RESPONSE TIME AND SATISFACTION BETWEEN DIFFERENT MODELS

Indicator Name	ADGN Model	CDTG Model	ST-GCNModel	GCSTModel
Response time (ms)	91	86	156	137
Satisfaction (%)	90	92	76	78
Feasibility (%)	91	93	84	81

Table III shows the response time and satisfaction between different models. The response times of the ADGN model and CDTG model proposed in the article were 91ms and 86ms, respectively. Compared with the ST-GCN model and GCST model, they had faster reaction speed and better low latency. Meanwhile, the above two models achieved 90% and 92% user satisfaction during the application process, and the feasibility under user evaluation also reached 91% and 93%, respectively. Therefore, the ADGN model and CDTG model proposed in this study had relatively superior practical application value in different regions.

V. DISCUSSION

Skeleton AR can be used in HAR to promote the expression generalization ability of human behavior. At present, research on skeleton AR has received widespread development and attention. Fang Z et al. used GCN to capture the activity status of skeleton joints and identify their movements. The accuracy of the skeleton AR model constructed in this experiment was only 87.79% [19]. Considering the background complexity of skeleton AR, this study optimized the skeleton AR model using AM and CDGN, and the final recognition accuracy of the models was above 90%. To improve the robustness of the skeleton AR model, Liu Y et al. used the KA-AGTN algorithm to construct a skeleton AR model mainly based on time series. Although it improved the recognition accuracy by 1.9% compared to traditional GCN models, it was lower than the model recognition accuracy proposed in this study [20]. The reason why the model proposed in the study was more excellent was that AM could enhance the model's ability to capture key actions, while CDGN could reduce noise interference in AR. Therefore, the optimized skeleton AR model had higher accuracy.

VI. CONCLUSION

To accurately capture and recognize spatiotemporal information and human actions in AR models, an ADGN

recognition model based on spatiotemporal CNN was proposed. By introducing the AM, the ADGN model enhanced the ability to obtain joint information. By integrating the central differential and spatiotemporal transformation networks simultaneously, a CDTG model could be constructed to compensate for the insufficient local dependency of joint information in the ADGN model. The experimental results showed that the ADGN model constructed in this study had CS benchmark recognition accuracy of 92.33%, 93.21%, and 94.12% in NTU RGB+D, MSR-Action 3D, and SBU datasets, respectively. The accuracy of CV recognition was 97.3%, 97.8%, and 96.7%, respectively. And the recognition accuracy of the ADGN model has reached 90%. The CDTG model had the highest recognition accuracy, with recognition accuracy rates of 92.87% and 97.52% in CS and CV, respectively. The response times of ADGN and CDTG models in application were 91ms and 86ms, respectively, with satisfaction rates of 90% and 92%, and feasibility rates of 91% and 93%, respectively. Therefore, ADGN and CDTG models proposed in this study had relatively superior practical application value in different regions. However, due to the small sample size of the entity and the fact that the model has not yet been AR validated in complex backgrounds, there is still room for improvement. At the same time, the combination of RGB or depth information is beneficial for optimizing the performance of the model, so other parameter conditions can be used to optimize the research model in the future.

VII. FUTURE RESEARCH WORK

Due to the fact that this study was not tested in a complex background, the test results of the model's performance were idealized. Considering the practical application of the model, the study will set up different complex environmental scenarios for model performance evaluation in subsequent experiments. The performance results were summarized and analyzed in complex scenarios, and the model is adjusted and optimized by setting different parameter conditions to improve its accuracy and efficiency. In addition, because the algorithm used in this study is an AM to enhance the performance of ADGN and CDTN models, and multimodal fusion forms such as RGB or depth information can also improve model performance. As a result, model performance tests can be conducted under different modalities of fusion to select the optimal multimodal fusion method to construct skeleton AR algorithms.

REFERENCES

- [1] H. Zhao, W. Xue, X. Li, Z. Gu, L. Niu, and L. Zhang, "Multi-mode neural network for human action recognition," *IET Comput. Vis.*, vol. 14, no. 8, pp. 587-596, Nov. 2020.
- [2] C. Peng, H. Huang, A. C. Tsoi, S. L. Lo, Y. Liu, and Z. Yang, "Motion boundary emphasised optical flow method for human action recognition," *IET Comput. Vis.*, vol. 14, no. 6, pp. 378-390, Jul. 2020.
- [3] W. R. Ko, M. Jang, J. Lee, and J. Kim, "AIR-Act2Act: Human-human interaction dataset for teaching non-verbal social behaviors to robots," *Int. J. Robot. Res.*, vol. 40, no. 4-5, pp. 691-697, Jan. 2021.
- [4] Y. Yang and X. Song, "Research on face intelligent perception technology integrating deep learning under different illumination intensities," *J. Comput. Cogn. Eng.*, vol. 1, no. 1, pp. 32-36, May. 2022.
- [5] Y. Lei, "Research on microvideo character perception and recognition based on target detection technology," *J. Comput. Cogn. Eng.*, vol. 1, no. 2, pp. 83-87, May. 2022.

- [6] B. Gu, W. Xiong, and Z. Bai, "Human action recognition based on supervised class-specific dictionary learning with deep convolutional neural network features," *Comput., Mater. Contin.*, vol. 63, no. 1, pp. 243-262, Mar. 2020.
- [7] H. Huang, H. Su, Z. Chang, M. Yu, J. Gao, X. Li, and S. Zheng, "Convolutional neural network with adaptive inferential framework for skeleton-based action recognition," *J. Vis. Commun. Image Represent.*, vol. 73, no. 11, pp. 102925.3-102925.10, Nov. 2020.
- [8] Z. Li, "Three-dimensional diffusion model in sports dance video human skeleton detection and extraction," *Adv. in Math. Phys.*, vol. 2021, no. 3, pp. 3772358.5-3772358.15, Sept. 2021.
- [9] C. Peng, H. Huang, A. Tsoi, S. Lo, Y. Liu, and Z. Yang, "Motion boundary emphasised optical flow method for human action recognition," *IET Comput. Vis.*, vol. 14, no.6, pp. 378-390, Jul. 2020.
- [10] J. Yang, "Study of human motion recognition algorithm based on multichannel 3D convolutional neural network," *Complex.*, vol. 2021, no. 18, pp. 7646813.54-7646813.62, May. 2021.
- [11] P. Zhao, D. Zhao, and X. Chen, "Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework," *IET Image Process.*, vol. 14, no. 7, pp. 1257-1264, Apr. 2020.
- [12] J. Liu and Y. Che, "Action recognition for sports video analysis using part-attention spatio-temporal graph convolutional network," *J. Electron. Imaging*, vol. 30, no. 3, pp. 33017.3-33017.16, Jun. 2021.
- [13] J. He, X. Wu, Z. Cheng, Z. Yuan, and Y. Jiang, "DB-LSTM: Densely-connected Bi-directional LSTM for human action recognition," *Neurocomputing*, vol. 444, no. 15, pp. 319-331, Jul. 2021.
- [14] W. Feng, Z. Hong, L. Wu, M. Fu, Y. Li, and P. Lin, "Network protocol recognition based on convolutional neural network," *China Commun.*, vol. 17, np. 4, pp. 125-139, Apr. 2020.
- [15] R. Ma, Z. Zhang, and E. Chen, "Human motion gesture recognition based on computer vision," *Complex.*, vol. 21, no. 5, pp. 6679746.15-6679746.26, Feb. 2021.
- [16] P. Chen, S. Guo, H. Li, X. Wang, G. Cui, C. Jiang, and L. Kong, "Through-wall human motion recognition based on transfer learning and ensemble learning," *IEEE Geosc. Remote Sens. Lett.*, vol. 191, no. 5, pp. 66-74, Apr. 2022.
- [17] H. Jiang and S. Tsai, "An empirical study on sports combination training action recognition based on smo algorithm optimization model and artificial intelligence," *Math. Probl. Eng.: Theory, Meth. Appl.*, vol. 2021, no. 31, pp. 7217383.46-7217383.51, Jul. 2021.
- [18] Z. Chen, X. Chen, Y. Ma, S. Guo, Y. Qin, and M. Liao, "Human posture tracking with flexible sensors for motion recognition," *Comput. Animat. Virtual Worlds*, vol. 32, no. 5, pp. 1993.56-1993.63, Apr. 2021.
- [19] Z. Fang, X. Zhang, T. Cao, Y. Zheng, and M. Sun, "Spatial-temporal slowfast graph convolutional network for skeleton-based action recognition," *IET Comput. Vis.*, vol. 16, no. 3, pp. 205-217, Nov. 2021.
- [20] Y. Liu, H. Zhang, D. Xu, and K. He, "Graph transformer network with temporal kernel attention for skeleton-based action recognition," *Knowl.-B. Syst.*, vol. 340, no. 15, pp. 1-16, Mar. 2022.

A Population-based Plagiarism Detection using DistilBERT-Generated Word Embedding

Yuqin JING*, Ying LIU

College of Electronical and Information Engineering, Chongqing Open University
Chongqing 400052, China

Abstract—Plagiarism is the unacknowledged use of another person’s language, information, or writing without crediting the source. This manuscript presents an innovative method for detecting plagiarism utilizing attention mechanism-based LSTM and the DistilBERT model, enhanced by an enriched differential evolution (DE) algorithm for pre-training and a focal loss function for training. DistilBERT reduces BERT’s size by 40% while maintaining 97% of its language comprehension abilities and being 60% quicker. Current algorithms utilize positive-negative pairs to train a two-class classifier that detects plagiarism. A positive pair consists of a source sentence and a suspicious sentence, while a negative pair comprises two dissimilar sentences. Negative pairs typically outnumber positive pairs, leading to imbalanced classification and significantly lower system performance. To combat this, a training method based on a focal loss (FL) is suggested, which carefully learns minority class examples. Another addressed issue is the training phase, which typically uses gradient-based methods like back-propagation for the learning process. As a result, the training phase has limitations, such as initialization sensitivity. A new DE algorithm is proposed to initiate the back-propagation process by employing a mutation operator based on clustering. A successful cluster for the current DE population is found, and a fresh updating approach is used to produce potential solutions. The proposed method is assessed using three datasets: SNLI, MSRP, and SemEval2014. The model attains excellent results that outperform other deep models, conventional, and population-based models. Ablation studies excluding the proposed DE and focal loss from the model confirm the independent positive incremental impact of these components on model performance.

Keywords—Plagiarism detection; LSTM; imbalanced classification; DistilBERT; differential evolution; focal loss

I. INTRODUCTION

With abundant information available online and powerful search engines, plagiarism has become a sensitive issue in various domains, including education. Plagiarism usually occurs intentionally or unknowingly [1]. In contrast, plagiarism techniques have practical uses in fields beyond detecting copied content, including retrieval of information [2] where some text is given as input and the most relevant matches returned.

Various techniques have been proposed in academic publications to address the challenge of detecting plagiarism. One prominent approach is the use of text distance methods, which aim to quantify the semantic proximity between two textual pieces by measuring the distance between them. Typically, there are three categories of text distances: length

distance, distribution distance, and semantic distance [3]. Length distance methods assess the resemblance between two texts by considering their numerical attributes. Popular techniques in this category include Euclidean distance, cosine distance, and Manhattan distance [4]. These methods rely on the numerical characteristics of the texts to calculate the degree of similarity. However, distance-based methods encounter two notable limitations. Firstly, they are often suitable only for symmetrical problems, which may restrict their applicability in certain scenarios. Additionally, using distance measures without considering the statistical characteristics of the data can be risky, particularly in cases such as question answering [5]. Distribution distances, on the other hand, offer an alternative approach to estimating the semantic similarity between two items by comparing their distributions. Techniques like Jensen–Shannon divergence [6] and Kullback–Leibler divergence [7] are commonly used in this category. These methods examine the lexical and semantic similarities between texts by analyzing the distributions of words or other linguistic features. By capturing the statistical properties of the data, distribution distances provide a more nuanced and comprehensive understanding of the semantic relationship between textual items. By leveraging distribution distances, researchers can effectively assess the similarity or dissimilarity between texts based on their underlying linguistic characteristics. These approaches take into account the broader context and semantic information, contributing to more accurate plagiarism detection.

Deep learning approaches have emerged as a powerful alternative to earlier methods in various fields, thanks to their inherent advantages, such as automated feature extraction [8]. Researchers have explored different deep learning architectures and techniques to tackle the task of sentence or text similarity and representation. One approach presented in [9] involves using a recurrent neural network (RNN) with word embeddings obtained from GloVe [10]. The RNN processes the words within a sentence and generates a representation of the sentence. Cosine distance metric [11] is then applied to measure the similarity between the sentence representations. In [40], a Siamese convolutional neural network (CNN) is introduced to capture the contextual information of individual words within a sentence. This network simultaneously produces a representation of word significance and the surrounding terms. By considering the local context, this approach aims to enhance the understanding of sentence meaning. Another RNN-based approach is presented in [12], where the textual data from corresponding words between sentence pairs is combined to create an internal representation.

This enables the model to capture the relationship between words in different sentences, contributing to a more comprehensive understanding of semantic similarity. In [13], a Long Short-Term Memory (LSTM) network is employed to extract high-level semantic information and measure the textual similarity between two sentences. The LSTM takes unprocessed pairs of sentence and word representations as input, allowing it to capture the complex semantic relationships within sentences. Attention-based models are also utilized in the pursuit of sentence similarity. In [14], an attention-based Siamese network is employed to determine the degree of similarity in meaning among sentences. The attention mechanism enables the model to focus on important elements within the sentences, enhancing its ability to capture semantic nuances. In [15], two different methods for answer selection based on similarity are introduced. One method incorporates a single transformer encoder along with embeddings from language models such as ELMo [16] and BERT [17]. The other method utilizes two pre-trained transformer encoders to capture the semantic information. Furthermore, [18] introduces the use of two Bidirectional LSTM (BLSTM) networks to independently derive sentence embeddings. Additionally, a revised data augmentation and loss function technique is implemented to address the challenge of imbalanced data distribution, which commonly occurs in sentence similarity tasks. One significant problem is the handling of imbalanced data distribution in plagiarism detection. The current algorithms often train two-class classifiers using positive-negative pairs, where negative pairs outnumber positive pairs. This imbalance negatively impacts system performance. Another limitation pertains to the training phase, which heavily relies on gradient-based methods like back-propagation. Although widely used, these methods have their own limitations, such as initialization sensitivity.

The unequal distribution of positive (plagiarized) and negative (non-plagiarized) cases pose a major obstacle in plagiarism detection. Failing to tackle this issue can result in a notable decline in performance. Approaches to tackle imbalanced class distribution can be categorized into two main types: the methods of the algorithm level and the data level. Data-level approaches aim to rectify the imbalanced distribution of classes by leveraging techniques such as over-sampling and under-sampling. One approach to address class imbalance is through the use of techniques such as the Synthetic Minority Oversampling Technique (SMOTE) [19], which creates instances by interpolating between adjacent minority examples. Another technique, NearMiss [20], involves under-sampling majority examples using the nearest neighbor algorithm. Over-sampling approaches might result in an overfitting problem, whereas applying under-sampling methods might lead to losing some helpful information about the dominant class. Algorithmic methods amplify the influence of the minority class based on techniques like ensemble learning [21], cost-sensitive learning [22], and decision threshold adjustment [23]. In the cost-sensitive approaches, various costs are assigned for misclassifications of different classes (higher costs for minority samples). The classification issue is framed as an optimization problem that seeks to minimize the total cost. Ensemble techniques train multiple classifiers and fuse the obtained results to reach a final

decision. Threshold adjustment approaches involve training a classifier and then modifying the threshold for classification during testing. Imbalanced classification has also been addressed using deep learning techniques [22, 24]. The study [25] develops a method to learn distinguishing features in unbalanced data while preserving inter-cluster and interclass margins. In [26] the author proposes a strategy that bootstraps convolutional network data into balance for each mini-batch.

Neural network methods, including deep networks, are usually based on gradient-based methods, including back-propagation, to find the appropriate network weights. Regrettably, these techniques are prone to be influenced by the initialization of parameters and may converge to suboptimal solutions. The quality of a neural network can be more significantly influenced by the initial weights than by the network structure and training samples [27]. Meta-heuristic algorithms, including differential evolution (DE) [28], have been proposed as a solution to address these issues and have demonstrated their effectiveness in optimizing the performance of the model [29, 30].

Differential Evolution (DE) is a robust method successfully utilized in various optimization tasks [31, 32]. It comprises three primary steps: mutation to create an additional candidate solution using scaling differences between solutions, crossover to integrate the produced solution with the initial solution, and selection to select the optimal solution for the subsequent iteration. The mutation operator is particularly important [33].

This article describes an original approach to plagiarism detection that employs a DE algorithm and attention-based LSTM model. The proposed model contains a feed-forward network to estimate the similarity degree between sentences and two LSTMs for source and suspicious sentences. The model is trained using pairs of sentences, including positive pairs with two similar sentences and negative pairs with two dissimilar sentences. DistilBERT word embedding is utilized, which can reduce BERT's size by 40% while maintaining 97% of its language comprehension abilities and being 60% quicker. The proposed DE algorithm utilizes clustering for weight initialization, aiming to detect an area in the exploration domain suitable for initiating the back-propagation (BP) algorithm. The best-performing solution from the top-performing cluster is selected as the starting point for the mutation operator, and a new approach for generating potential solutions is employed. Additionally, the proposed algorithm incorporates FL to address class imbalance. The model is assessed on SNLI, MSRP, and SemEval2014 datasets, demonstrating superior performance compared to other methods.

The main contributions of the article are as follows: 1) The article introduces a new DE algorithm that initiates the back-propagation process by employing a mutation operator based on clustering. This approach helps overcome limitations associated with initialization sensitivity, which is a common issue in gradient-based methods used during the training phase, 2) The article addresses the challenge of imbalanced class distribution in plagiarism detection, where negative pairs outnumber positive pairs. The proposed training method based on focal loss enables the model to better learn from minority

class examples, leading to improved system performance, 3) The article introduces the DistilBERT model that can reduce the size of BERT. This reduction in size leads to improved efficiency, making the model 60% quicker compared to the original BERT model, and 4) Ablation studies are conducted to evaluate the individual contributions of the DE algorithm and focal loss. The results confirm that these components have an independent positive incremental impact on the model's performance.

The article's residual parts are structured as follows. Section II provides a number of contextual information, whereas Section III outlines the proposed method for identifying plagiarism. Section IV gives the prediction of the study made. In Section V, the results of the experiments are presented, and Section VI summarizes the paper.

II. DIFFERENTIAL EVOLUTION

Differential evolution [28] has effectively optimized various problems [35, 36]. DE starts with an initial population, usually drawn from a random distribution, and comprises three primary operations: mutation, crossover, and selection. The mutation operation generates a mutant vector as

$$\vec{v}_{i,g} = \vec{x}_{r_1,g} + F(\vec{x}_{r_2,g} - \vec{x}_{r_3,g}) \quad (1)$$

where $\vec{x}_{r_1,g}$, $\vec{x}_{r_2,g}$ and \vec{x}_{r_3} are three randomly chosen candidate solutions from the available population, and F shows a factor scaling.

Crossover incorporates the mutant and target vectors. A well-known crossover operator is a binomial crossover, which does this as

$$u_{i,j,g} = \begin{cases} v_{i,j,g} & \text{if } \text{rand}(0,1) \leq CR \text{ or } j = j_{rand} \\ x_{i,j,g} & \text{otherwise} \end{cases} \quad (2)$$

Where CR denotes the rate of crossover, and j_{rand} is a number chosen randomly from the set $\{1,2,\dots,D\}$, where D is the dimensionality of a candidate solution.

Lastly, the selection operator elects the superior solution from the target and trial vectors.

III. PROPOSED APPROACH

The overall structure of the suggested method is displayed in Fig. 1. As seen, it comprises three main stages, pre-processing, word embedding, and prediction. First, redundant words and symbols are removed from the sentences. Next, the embedding vector of each word is obtained using BERT, and ultimately, the model predicts the similarity between the two sentences. The proposed model incorporates a clustering-based differential evolution algorithm to find the initial seeds of the network weights while using focal loss to handle class imbalance.

A. Pre-Processing

Data pre-processing is a crucial aspect of any NLP system as the fundamental characters, words, and sentences extracted in this phase are forwarded to the subsequent stages. Consequently, they considerably impact the outcome. Conversely, an unsuitable pre-processing technique can decrease the model's performance [37]. Common stop-word elimination and stemming techniques are used in the approach.

Stop words are part of sentences that can be regarded as overhead. The most common stop words are articles, prepositions, pronouns, etc. They should thus be removed as they cannot function as keywords in text mining applications [38] and decrease the number of dimensions in the term space

Stemming is employed to determine the base form of a word. For instance, the terms 'watch', 'watched', 'watching', 'watcher', etc., can all be reduced to the stem word "watch" by stemming. Stemming reduces ambiguity, decreases the number of words, and minimizes time and memory requirements [37].

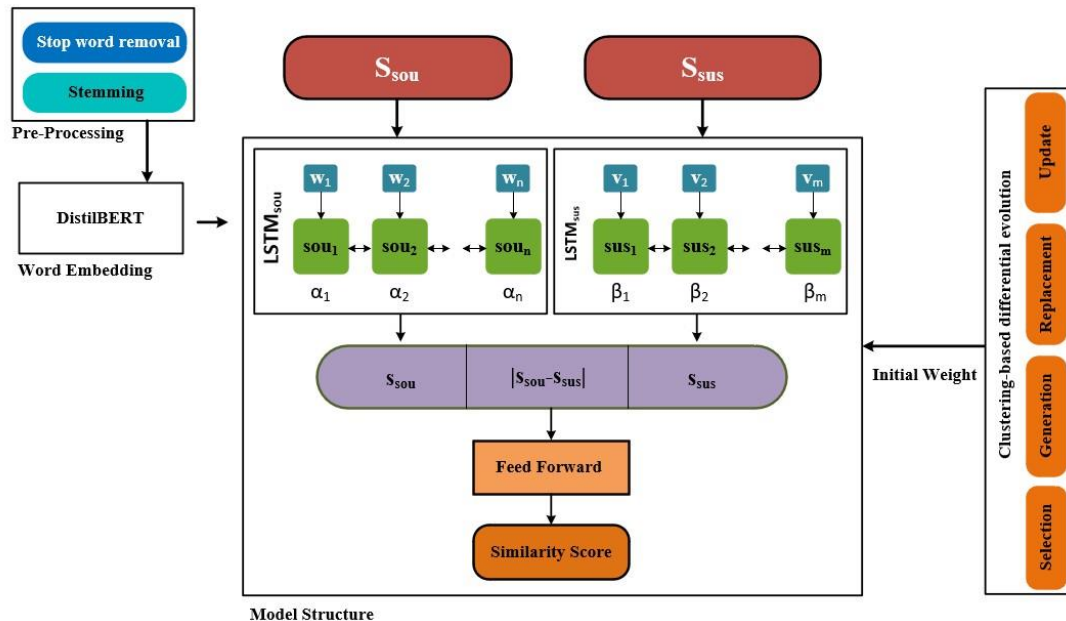


Fig. 1. Architecture of the suggested model as a whole.

IV. PREDICTION

The prediction model comprises two attention-based LSTM networks as extractors of the embeddings of the source and suspicious sentences and a feed-forward network as a predictor of the similarity of the two sentences. Given $s_{sus} = \{v_1, v_2, \dots, v_m\}$ and $s_{sou} = \{w_1, w_2, \dots, w_n\}$ as the sentence of the suspicious and source, where v_i and w_i denote the i -th word in the suspicious and source sentences, respectively. s_{sou} and s_{sus} are restricted to n and m words due to the length limitation in BLSTM (in the work, $n = m$). s_{sou} and s_{sus} are nourished separately into an LSTM network. These sentences embeddings are computed using the mechanism of attention, as

$$s_{sou} = \sum_{i=1}^n \alpha_i h_{sou_i} \quad (3)$$

And

$$s_{sus} = \sum_{i=1}^m \beta_i h_{sus_i} \quad (4)$$

where the i -th hidden vectors are represented in the BLSTM by $h_{sou_i} = [\vec{h}_{sou_i}, \vec{h}_{sou_i}]$ and $h_{sus_i} = [\vec{h}_{sus_i}, \vec{h}_{sus_i}]$, and the i -th attention weight for every section is shown in the BLSTM by $\alpha_i, \beta_i \in [0,1]$, computed as

$$\alpha_i = \frac{e^{u_i}}{\sum_{j=1}^n e^{u_j}} \quad (5)$$

And

$$\beta_i = \frac{e^{v_i}}{\sum_{j=1}^m e^{v_j}} \quad (6)$$

With

$$u_i = \tanh(W_u h_{sou_i} + b_u) \quad (7)$$

And

$$v_i = \tanh(W_v h_{sus_i} + b_v) \quad (8)$$

where W_u, W_v, b_u and b_v are the weight matrices and biases to the attention mechanisms. The fully-connected network's input is the connection of the s_{sou}, s_{sus} and $|s_{sou} - s_{sus}|$ as shown in Fig. 1. The dataset used for training consists of positive and negative pairs, where positive pairs contain a source sentence and a copied sentence and negative pairs comprise a source sentence and a different sentence.

The model has two training phases, pre-training and fine-tuning. In pre-training, an appropriate starting configuration is found. The weights obtained in pre-training are then the initial weights of the fine-tuning phase. In the pre-training phase, the enhanced differential evolution algorithm is employed.

A. Pre-Training

At this stage, the weights of the LSTM, attention mechanism, and feed-forward neural network are initialized. For this, an enhanced differential evolution method is introduced, boosted by a clustering scheme and a novel fitness function.

1) *Clustering-based differential evolution*: A clustering-based mutation and updating scheme is employed in the enhanced DE algorithm to improve the optimization performance.

The suggested mutation operator, which takes inspiration from [39] pinpoints a propitious area in the search space. The k -means clustering technique is used to partition the current population P into k clusters, each defining a distinct section of the search space. From [2, N], a random integer is chosen to depict the clusters number. The cluster with the lowest mean fitness of its samples is the best cluster after clustering.

The suggested mutation based on clustering is described as follows:

$$\vec{v}^{clu}_i = \overline{win}_g + F(\vec{x}_{r_1,g} - \vec{x}_{r_2,g}) \quad (9)$$

where \overline{win}_g is the most acceptable solution in the promising region, and $\vec{x}_{r_1,g}$ and $\vec{x}_{r_2,g}$ are two randomly determined solutions from the available population. It should be noted that wing is not always the population's most acceptable solution. The procedure of the mutation on the basis of the clustering is implemented M times.

When M new solutions have been provoked through clustering-based mutation, the current population is updated. The steps are as follows:

- Selection: Generate k individuals randomly as the starting points of k -means;
- Generation: Generate the solutions of the M by applying clustering-based mutation as the collection v^{clu} ;
- Replacement: Choose M solutions at random and determine as B ;
- Update: The best M solutions from the $v^{clu} \cup B$ determined as the B' . The novel population is afterwards calculated as $(P - B) \cup B'$.

2) *Encoding strategy*: The primary structure of the proposed model includes two LSTM networks along with their attention mechanisms and a feed-forward network. As illustrated in Fig. 2, all weights and bias terms are arranged into a vector to form a candidate solution in the proposed DE algorithm.

3) *Fitness function*: To calculate the quality of a candidate solution, the fitness function is as

$$Fitness = \frac{1}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (10)$$

Where N is the number of training examples, y_i and \hat{y}_i show the i -th target and output predicted by the model, respectively.

B. Focal Loss

The plagiarism problem is defined as a two-class classification problem based on positive and negative classes. As an imbalanced problem, with few samples in the negative class, focal loss (FL) [34] is used to address this.

FL is a modification of binary cross-entropy (CE) that focuses training on harder (i.e., minority class) samples [40]. CE is defined as

$$CE = \begin{cases} -\log(p), & y = 1 \\ -\log(1-p), & \text{otherwise} \end{cases} \quad (11)$$

where $y \in \{-1,1\}$ is the actual class label, and $p \in [0,1]$ is the predicted probability of the model for the class with target $y = 1$. The probability is

$$p_t = \begin{cases} -p, & y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (12)$$

and hence

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (13)$$

FL tends to add a modulating component to cross-entropy loss, leading to

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (14)$$

Where $\gamma > 0$ (if $\gamma = 1$, then FL is similar to CE loss), and $\alpha \in [0,1]$ is the inverse class frequency.

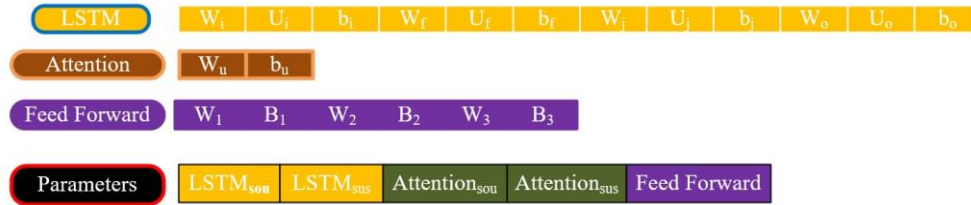


Fig. 2. Encoding strategy in the proposed algorithm.

V. RESULTS

A. Datasets

In the tests, the following three benchmark datasets are utilized:

- SNLI: the Stanford Natural Language Inference (SNLI) corpus [41] is a large dataset consisting of pairs of labelled sentences with three classes, including contradiction, entailment, and semantic independence. It comprises 550,152 sentence pairs for training and 10,000 pairs of sentences each for testing and validation.
- MSRP: A rephrasing set of data from Internet news articles called the Microsoft Research Paraphrase Corpus (MSRP) [42], divided into training and testing, with pairs of positive and negative sentences by several experts. Of the whole collection, about 67% of paraphrases are present. The test and training datasets have 1,726 and 4,076 examples, respectively, out of which 1,147 and 2,753 are paraphrases, respectively.
- SemEval2014: the Semantic Evaluation Database (SemEval) [13] is a widely-used benchmark for evaluating STS, presented in various versions. The Compositional Knowledge (SICK) dataset [43] from 2014 is employed to assess the semantic similarity of

sentences. The dataset includes 10,000 sentence pairs, distributed as 4,500 pairs for training, 500 for validation, and 5,000 for testing.

B. Model Performance

The algorithm is compared to seven deep learning methods, namely RNN [9], Siamese CNN+LSTM [44], CA-RNN [14], AttSiaBiLSTM [13], LSTM+FNN+attention [14], CETE [15] and STS-AM [18]. The results are given in Tables I, II and III for SNLI, MSRP, and SemEval2014, correspondingly. For the suggested approach, outcomes are presented based on accidental weight initialization, the use of FL, and the full proposed model. The proposed model demonstrates superior performance compared to other models, including CETE, the best-performing competitor, across all metrics for SNLI. The error rate is reduced by over 50% and 54% in the two primary metrics, F-measure and G-means. Comparing the proposed model with Proposed+random weights and Proposed+random weights+FL, the mistake percentage is reduced by around 67%, highlighting the significance of improved DE and FL methodologies. The proposed model achieved the most significant improvement for the MSRP dataset, followed by the CETE algorithm. The error rate improvement for this dataset is about 27.41% and 26.69% for both the F-measure and G-means criteria, respectively. In the SemEval2014 dataset, the proposed method reduces the classification mistake by over 18% and 37% compared to CETE and STS-AM, respectively.

TABLE I. COMPARATIVE PERFORMANCE OF DEEP LEARNING MODELS ON THE SNLI DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY $10^{(-3)}$, FOR CONCISE PRESENTATION

Model	Accuracy	Recall	Precision	F-measure	G-means
RNN [9]	687×10^{-3}	594×10^{-3}	540×10^{-3}	566×10^{-3}	661×10^{-3}
Siamese CNN+LSTM [44]	850×10^{-3}	763×10^{-3}	792×10^{-3}	777×10^{-3}	826×10^{-3}
CA-RNN [14]	790×10^{-3}	667×10^{-3}	704×10^{-3}	685×10^{-3}	754×10^{-3}
AttSiaBiLSTM [13]	695×10^{-3}	569×10^{-3}	554×10^{-3}	561×10^{-3}	658×10^{-3}
LSTM+FNN+attention [14]	818×10^{-3}	781×10^{-3}	715×10^{-3}	747×10^{-3}	809×10^{-3}
CETE [15]	874×10^{-3}	855×10^{-3}	795×10^{-3}	824×10^{-3}	870×10^{-3}
STS-AM [18]	756×10^{-3}	625×10^{-3}	650×10^{-3}	637×10^{-3}	718×10^{-3}
Proposed+random weights	808×10^{-3}	777×10^{-3}	698×	735×	801×
Proposed+random weights+FL	815×10^{-3}	784×10^{-3}	708×	744×	808×
Proposed	930×10^{-3}	920×10^{-3}	881×	900×	927×

TABLE II. COMPARATIVE PERFORMANCE OF DEEP LEARNING MODELS ON THE MSRP DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Model	Accuracy	Recall	Precision	F-measure	G-means
RNN [9]	853×10^{-3}	922×10^{-3}	866×10^{-3}	893×10^{-3}	812×10^{-3}
Siamese CNN+LSTM [44]	863×10^{-3}	916×10^{-3}	882×10^{-3}	899×10^{-3}	833×10^{-3}
CA-RNN [14]	880×10^{-3}	928×10^{-3}	896×10^{-3}	912×10^{-3}	854×10^{-3}
AttSiaBiLSTM [13]	874×10^{-3}	927×10^{-3}	889×10^{-3}	908×10^{-3}	845×10^{-3}
LSTM+FNN+attention [14]	889×10^{-3}	917×10^{-3}	916×10^{-3}	916×10^{-3}	873×10^{-3}
CETE [15]	916×10^{-3}	949×10^{-3}	926×10^{-3}	937×10^{-3}	898×10^{-3}
STS-AM [18]	899×10^{-3}	940×10^{-3}	910×10^{-3}	925×10^{-3}	876×10^{-3}
Proposed+random weights	875×10^{-3}	908×10^{-3}	905×10^{-3}	906×10^{-3}	858×10^{-3}
Proposed+randomweights+FL	895×10^{-3}	926×10^{-3}	917×10^{-3}	921×10^{-3}	879×10^{-3}
Proposed	937×10^{-3}	961×10^{-3}	946×10^{-3}	953×10^{-3}	925×10^{-3}

TABLE III. COMPARATIVE PERFORMANCE OF DEEP LEARNING MODELS ON THE SEMEVAL2014 DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Model	Accuracy	Recall	Precision	F-measure	G-means
RNN [9]	809×10^{-3}	822×10^{-3}	963×10^{-3}	887×10^{-3}	750×10^{-3}
Siamese CNN+LSTM [44]	775×10^{-3}	787×10^{-3}	958×10^{-3}	864×10^{-3}	720×10^{-3}
CA-RNN [14]	811×10^{-3}	826×10^{-3}	961×10^{-3}	888×10^{-3}	742×10^{-3}
AttSiaBiLSTM [13]	799×10^{-3}	816×10^{-3}	957×10^{-3}	881×10^{-3}	720×10^{-3}
LSTM+FNN+attention [14]	733×10^{-3}	746×10^{-3}	949×10^{-3}	835×10^{-3}	670×10^{-3}
CETE [15]	854×10^{-3}	868×10^{-3}	969×10^{-3}	916×10^{-3}	791×10^{-3}
STS-AM [18]	823×10^{-3}	834×10^{-3}	966×10^{-3}	895×10^{-3}	768×10^{-3}
Proposed +random weights	839×10^{-3}	855×10^{-3}	964×10^{-3}	906×10^{-3}	766×10^{-3}
Proposed +random weights+FL	849×10^{-3}	863×10^{-3}	967×10^{-3}	912×10^{-3}	783×10^{-3}
Proposed	876×10^{-3}	884×10^{-3}	977×10^{-3}	928×10^{-3}	838×10^{-3}

C. Comparison with other Metaheuristics

The enhanced DE algorithm is contrasted with several metaheuristic optimization algorithms in the subsequent experiment. Different metaheuristics are used to obtain the initial model parameters while keeping the same as the other model components, i.e., pre-processing, word embedding, LSTM and network structure, and loss function. Eight different algorithms, namely (standard) DE [45], FA [46], BA [47], COA [48], ABC [49], GWO [50], WOA [51], and SSA [52], are used. The obtained results are reported in Tables IV, V and VI for the SNLI, MSRP, and SemEval2014 datasets, respectively. For the SNLI dataset, the suggested model reduces error by about 44% compared to the standard DE. It clearly shows that the proposed model has a substantial ability compared to the standard one. Also, DE offers more acceptable results than other algorithms, including ABC, GWO, and BAT. There is a minor improvement for the other two datasets, so the error rate for MSRP and SemEval2014 is reduced by around 19.17% and 8.82%, respectively.

D. Word Embeddings

Word embedding is a crucial component of In-depth learning-based models since the input is read as a vector, and if the embedding is erroneous, the model might be misled. This study used the DistilBERT model as a word embedding, one of the most recent embedding models. Five more-word embeddings are used to compare various word embeddings to the model: One-Hot encoding One-Hot encoding [53], CBOW, Skip-gram [54], GloVe [10], and FastText [55]. One-Hot

encoding is a crucial step in changing the collected data variables fed to In-depth learning methods, enhancing the accuracy of predictions and classifications. It generates a binary feature for every class, and each sample's feature is given a value of 1 corresponding to its original class. Skip-gram and CBOW are techniques that transform a word into its corresponding representation vector using neural networks. The GloVe is a method for aggregating global word-word co-occurrence data from a corpus. The Skip-gram paradigm is expanded by the word embedding technique known as FastText. This approach encodes each word as an n-gram of letters rather than learning word vectors. The outcomes of this experiment can be found in Tables VII, VIII and IX for the SNLI, MSRP, and SemEval2014 datasets, respectively. The worst-performing word embedding method was One Hot encoding. In the MSRP dataset, the proposed model showed an improvement of approximately 85.81% and 83.51% for the two criteria, F-measure and G-means, respectively. Skip-gram and CBOW operate nearly similarly across the three datasets because of similar architecture, which is superior to the Glove model. FastText performs better than other models but poorly on BERT. The error rate is reduced by more than 18%, 15%, and 24% for the SNLI, MSRP, and SemEval2014 datasets, respectively, when utilizing BERT instead of FastText.

E. Loss Functions

Finally, to justify the selection of focal loss in the approach, the comparison is made with four other loss functions, namely weighted cross-entropy (WCE) [56], balanced cross-entropy (BCE) [57], Dice loss (DL) [58], and Tversky loss (TL) [59].

The results of these experiments are given in Tables X, XI and XII for the SNLI, MSRP, and SemEval2014 datasets, respectively. The use of focal loss gives the best results for all measures on the SNLI and MSRP datasets and yields the best G-means results for all three datasets. The results of this experiment are given in Tables X, XI and XII for the SNLI, MSRP, and SemEval2014 datasets, respectively. Generally speaking, the reduction of FL error compared to TL for SNLI and MSRP datasets is about 19% and 27%. However, these two functions are slightly different in the SemEval2014 dataset, so the improvement rate for this dataset is about 12%.

F. Examples

A qualitative example is provided to demonstrate the important contributions of both the improved DE algorithm and the use of FL in the approach. The source sentence "Two people are kickboxing, and spectators are watching" from the SemEval2014 dataset is used for this purpose. Fig. 3 gives the results of the top five sentences retrieved by the BPD model with random weight initialization and focal loss, without FL, and the full approach. As is apparent, the full model extracts suspicious sentences most similar to the source sentence, while the other two models retrieve these only in the lower rankings.

TABLE IV. COMPARATIVE PERFORMANCE OF METAHEURISTIC ALGORITHMS ON THE SNLI DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY $10^{(-3)}$, FOR CONCISE PRESENTATION

Algorithm	Accuracy	Recall	Precision	F-measure	G-means
DE	897×10^{-3}	889×10^{-3}	824×10^{-3}	855×10^{-3}	895×10^{-3}
FA	864×10^{-3}	803×10^{-3}	801×10^{-3}	802×10^{-3}	848×10^{-3}
BA	876×10^{-3}	850×10^{-3}	801×10^{-3}	825×10^{-3}	870×10^{-3}
COA	860×10^{-3}	811×10^{-3}	787×10^{-3}	799×10^{-3}	847×10^{-3}
ABC	885×10^{-3}	869×10^{-3}	809×10^{-3}	838×10^{-3}	881×10^{-3}
GWO	842×10^{-3}	780×10^{-3}	763×10^{-3}	771×10^{-3}	826×10^{-3}
WOA	883×10^{-3}	832×10^{-3}	828×10^{-3}	830×10^{-3}	870×10^{-3}
SSA	863×10^{-3}	820×10^{-3}	789×10^{-3}	804×10^{-3}	852×10^{-3}

TABLE V. COMPARATIVE PERFORMANCE OF METAHEURISTIC ALGORITHMS ON THE MSRP DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY $10^{(-3)}$, FOR CONCISE PRESENTATION

Algorithm	Accuracy	Recall	Precision	F-measure	G-means
DE	925×10^{-3}	959×10^{-3}	930×10^{-3}	944×10^{-3}	906×10^{-3}
FA	897×10^{-3}	942×10^{-3}	908×10^{-3}	925×10^{-3}	874×10^{-3}
BA	910×10^{-3}	944×10^{-3}	922×10^{-3}	933×10^{-3}	892×10^{-3}
COA	902×10^{-3}	936×10^{-3}	918×10^{-3}	927×10^{-3}	884×10^{-3}
ABC	899×10^{-3}	946×10^{-3}	906×10^{-3}	926×10^{-3}	873×10^{-3}
GWO	884×10^{-3}	926×10^{-3}	902×10^{-3}	914×10^{-3}	861×10^{-3}
WOA	901×10^{-3}	938×10^{-3}	915×10^{-3}	926×10^{-3}	881×10^{-3}
SSA	890×10^{-3}	929×10^{-3}	908×10^{-3}	918×10^{-3}	869×10^{-3}

TABLE VI. COMPARATIVE PERFORMANCE OF METAHEURISTIC ALGORITHMS ON THE SEMEVAL2014 DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY $10^{(-3)}$, FOR CONCISE PRESENTATION

Algorithm	Accuracy	Recall	Precision	F-measure	G-means
DE	864×10^{-3}	873×10^{-3}	975×10^{-3}	921×10^{-3}	822×10^{-3}
FA	854×10^{-3}	865×10^{-3}	972×10^{-3}	915×10^{-3}	807×10^{-3}
BA	860×10^{-3}	869×10^{-3}	973×10^{-3}	918×10^{-3}	814×10^{-3}
COA	856×10^{-3}	867×10^{-3}	971×10^{-3}	916×10^{-3}	805×10^{-3}
ABC	851×10^{-3}	863×10^{-3}	970×10^{-3}	913×10^{-3}	796×10^{-3}
GWO	848×10^{-3}	857×10^{-3}	972×10^{-3}	911×10^{-3}	805×10^{-3}
WOA	844×10^{-3}	856×10^{-3}	969×10^{-3}	909×10^{-3}	788×10^{-3}
SSA	843×10^{-3}	857×10^{-3}	966×10^{-3}	908×10^{-3}	777×10^{-3}

TABLE VII. COMPARATIVE PERFORMANCE OF WORD EMBEDDINGS ON THE SNLI DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY $10^{(-3)}$, FOR CONCISE PRESENTATION

Word embedding	Accuracy	Recall	Precision	F-measure	G-means
One-Hot encoding	650×10^{-3}	473×10^{-3}	489×10^{-3}	481×10^{-3}	592×10^{-3}
CBOW	856×10^{-3}	779×10^{-3}	796×10^{-3}	787×10^{-3}	835×10^{-3}
Skip-gram	871×10^{-3}	817×10^{-3}	808×10^{-3}	812×10^{-3}	857×10^{-3}
GloVe	845×10^{-3}	798×10^{-3}	762×10^{-3}	780×10^{-3}	833×10^{-3}
FastText	905×10^{-3}	861×10^{-3}	861×10^{-3}	861×10^{-3}	893×10^{-3}
BERT	912×10^{-3}	892×10^{-3}	910×10^{-3}	886×10^{-3}	902×10^{-3}

TABLE VIII. COMPARATIVE PERFORMANCE OF WORD EMBEDDINGS ON THE MSRP DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Word embedding	Accuracy	Recall	Precision	F-measure	G-means
One-Hot encoding	604×10^{-3}	659×10^{-3}	721×10^{-3}	689×10^{-3}	571×10^{-3}
CBOW	802×10^{-3}	840×10^{-3}	859×10^{-3}	849×10^{-3}	781×10^{-3}
Skip-gram	830×10^{-3}	856×10^{-3}	884×10^{-3}	870×10^{-3}	816×10^{-3}
GloVe	781×10^{-3}	824×10^{-3}	844×10^{-3}	834×10^{-3}	758×10^{-3}
FastText	864×10^{-3}	880×10^{-3}	913×10^{-3}	896×10^{-3}	857×10^{-3}
BERT	912×10^{-3}	900×10^{-3}	922×10^{-3}	910×10^{-3}	913×10^{-3}

TABLE IX. COMPARATIVE PERFORMANCE OF WORD EMBEDDINGS ON THE SEMEVAL2014 DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Word embedding	Accuracy	Recall	Precision	F-measure	G-means
One-Hot encoding	504×10^{-3}	529×10^{-3}	875×10^{-3}	659×10^{-3}	367×10^{-3}
CBOW	749×10^{-3}	768×10^{-3}	946×10^{-3}	848×10^{-3}	659×10^{-3}
Skip-gram	758×10^{-3}	773×10^{-3}	951×10^{-3}	853×10^{-3}	686×10^{-3}
GloVe	697×10^{-3}	715×10^{-3}	936×10^{-3}	811×10^{-3}	606×10^{-3}
FastText	812×10^{-3}	826×10^{-3}	961×10^{-3}	888×10^{-3}	742×10^{-3}
BERT	842×10^{-3}	862×10^{-3}	970×10^{-3}	901×10^{-3}	763×10^{-3}

TABLE X. COMPARATIVE PERFORMANCE OF LOSS FUNCTION ON THE SNLI DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Loss function	Accuracy	Recall	Precision	F-measure	G-means
WCE	871×10^{-3}	856×10^{-3}	788×10^{-3}	821×10^{-3}	868×10^{-3}
BCE	895×10^{-3}	874×10^{-3}	828×10^{-3}	850×10^{-3}	890×10^{-3}
DL	915×10^{-3}	885×10^{-3}	870×10^{-3}	877×10^{-3}	908×10^{-3}
TL	905×10^{-3}	880×10^{-3}	848×10^{-3}	864×10^{-3}	899×10^{-3}

TABLE XI. COMPARATIVE PERFORMANCE OF LOSS FUNCTION ON THE MSRP DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Loss function	Accuracy	Recall	Precision	F-measure	G-means
WCE	861×10^{-3}	906×10^{-3}	887×10^{-3}	896×10^{-3}	836×10^{-3}
BCE	883×10^{-3}	923×10^{-3}	903×10^{-3}	913×10^{-3}	861×10^{-3}
DL	915×10^{-3}	943×10^{-3}	930×10^{-3}	936×10^{-3}	899×10^{-3}
TL	899×10^{-3}	933×10^{-3}	917×10^{-3}	925×10^{-3}	881×10^{-3}

TABLE XII. COMPARATIVE PERFORMANCE OF LOSS FUNCTION ON THE SEMEVAL2014 DATASET. THE PERFORMANCE METRICS ARE REPRESENTED IN FRACTIONS, MULTIPLIED BY 10^{-3} , FOR CONCISE PRESENTATION

Loss function	Accuracy	Recall	Precision	F-measure	G-means
WCE	876×10^{-3}	904×10^{-3}	957×10^{-3}	930×10^{-3}	736×10^{-3}
BCE	875×10^{-3}	899×10^{-3}	960×10^{-3}	928×10^{-3}	754×10^{-3}
DL	877×10^{-3}	890×10^{-3}	972×10^{-3}	929×10^{-3}	815×10^{-3}
TL	876×10^{-3}	895×10^{-3}	966×10^{-3}	929×10^{-3}	784×10^{-3}

rank	Proposed+random weights+RL	Proposed without RL	Proposed
1	Two people are wading through the water	Two people are riding a motorcycle	Two people are fighting and spectators are watching
2	A few men are watching cricket	Two people are fighting and spectators are watching	Two spectators are kickboxing and some people are watching
3	Two girls are laughing breathlessly and other girls are watching them	Two adults are sitting in the chairs and are watching the ocean	Two young women are sparring in a kickboxing fight
4	Two people wearing snowsuits are on the ground making snow angels	Two spectators are kickboxing and some people are watching	Two women are sparring in a kickboxing match
5	Two spectators are kickboxing and some people are watching	A few men are watching cricket	Two people are wading through the water

Fig. 3. Top-ranked suspicious sentences for source sentence “Two people are kickboxing, and spectators are watching.” Words that appear in the source sentence are bolded.

G. Discussion

The article presented an innovative approach to plagiarism detection by using an attention mechanism-based LSTM and the DistilBERT model. The utilization of the DistilBERT model is particularly notable as it reduces the size of the original BERT model by 40% while maintaining 97% of its language comprehension capabilities and increasing the speed by 60%. Two novel approaches were introduced to enhance the overall performance of the system. First, a focal loss function was used to address the issue of imbalanced classification, which often occurs when negative pairs significantly outnumber positive pairs. Second, an enriched DE algorithm was introduced to address the limitations associated with traditional gradient-based learning methods, such as initialization sensitivity. The approach was evaluated on three benchmark datasets: SNLI, MSRP, and SemEval2014, and it outperformed other deep models and conventional and population-based models. The effectiveness of the DE algorithm and the focal loss function was further validated through ablation studies.

While the study exhibits considerable promise, there are a few potential limitations:

- The study indeed leverages benchmark datasets that provide reliable standards for performance evaluation. They serve as a crucial starting point for developing and refining the model, and their use enables the results to be compared directly with other models that have also been evaluated using these datasets. However, these datasets, although comprehensive and widely used, may not fully encapsulate the full diversity and complexity of real-world plagiarism scenarios. Real-world plagiarism can be exceptionally intricate, involving subtle paraphrasing, strategic insertion of synonyms, reordering of sentences, or blending of original and copied material, among other tactics. These practices can often deceive conventional plagiarism detection tools, requiring models that can comprehend and identify such complex forms of plagiarism. Furthermore, these benchmark datasets might lack certain forms of plagiarism seen in specific fields or cultures. Plagiarism, after all, can differ greatly across different academic disciplines, professional fields, and cultural contexts. For instance, the plagiarism practices in a literature research paper could be entirely different from those in a technical report in engineering. Lastly, the real-world plagiarism scenarios are continually evolving, influenced by the advancement of technology and changes in writing and copying techniques. The dynamic and constantly changing nature of real-world plagiarism can present a challenge that these static, fixed datasets may not be fully equipped to address.
- The incorporation of DistilBERT is indeed a step forward in reducing the model's size and enhancing its operational speed, owing to its design that maintains substantial language comprehension capabilities while being considerably smaller and faster than the original BERT model. This makes the model more feasible for applications that demand quicker processing times and

limited memory capacities. However, even with these benefits, the combined system that also includes the attention mechanism-based LSTM and the enriched DE algorithm may still have significant computational demands. The LSTM component, known for its ability to remember long-term dependencies in sequence data, can be computationally intensive, particularly for longer sequences or larger datasets. The recurrent nature of LSTMs, where outputs from one step are fed as inputs to the next, makes parallelization of computations difficult, potentially slowing down the training process. On the other hand, the DE algorithm, while providing an innovative solution to the limitation of sensitivity to initialization inherent in gradient-based training methods, adds another layer of complexity to the system. The operations involved in differential evolution, such as mutation, recombination, and selection, while aiding the optimization process, also contribute to the computational burden. Moreover, the system must be trained on multiple iterations to effectively learn from the data, and each iteration involves processing the entire dataset. The computational demands can, therefore, escalate with the volume of data, length of sequences, and complexity of the tasks at hand. In the real world, these requirements could translate to higher memory and processing power requirements, extended training times, and increased energy consumption. They could also limit the system's deployability on devices with limited computational capabilities, such as mobile devices or low-end personal computers. Therefore, while the use of DistilBERT, LSTM, and the DE algorithm offers various advantages, further work could be directed towards optimizing the system to make it more efficient and less resource-intensive.

Finally, future works that can be considered are as follows:

- Investigating the model's performance on other languages beyond those in the current datasets could be beneficial, potentially leading to the development of a more universally applicable plagiarism detection tool.
- It would be valuable to test the model in real-world scenarios, such as academic papers or professional reports, to further assess its effectiveness and robustness.
- There may be scope to optimize the DE algorithm further for this specific use case. Tuning the parameters of the mutation operator based on the characteristics of the plagiarism detection task could potentially enhance the system's performance.
- Exploring the use of other pre-trained language models, including GPT-3 and T5, and compare their performance with DistilBERT. Comparing the capabilities of multiple pre-trained language models such as GPT-3 and T5 for plagiarism detection tasks could provide valuable insights into the suitability of these models for this application. GPT-3 is a transformer-based language model trained on a massive corpus of diverse texts and has shown impressive

results in a variety of natural language processing tasks. T5, on the other hand, is a text-to-text transformer that can be fine-tuned for different tasks, including text classification and sequence labeling.

VI. CONCLUSION

Plagiarism is the unacknowledged use of another individual's language, information, or writing without crediting the source. An innovative model was introduced to detect plagiarism based on DistilBERT word embeddings, an LSTM approach with an attention mechanism, and an enhanced DE algorithm used for pre-training the networks. To address the issue of inherent class imbalance, focal loss was employed. The enhanced DE algorithm groups the present population to pinpoint a potential area within the search space and integrates a novel update mechanism. DistilBERT can improve the performance of BERT by 40% and 97% in terms of size, language comprehension abilities, and speed, respectively. Extensive experiments on three datasets confirm the approach to yield excellent performance, outperforming various plagiarism detection approaches. The DE algorithm is superior to several other meta-heuristic methods. In forthcoming studies, the plan is to utilize the technique on several deep models, and an investigation of a version of the algorithm that can handle multiple objectives is underway.

ACKNOWLEDGMENT

This work was supported by the Science and Technology Research Program of the Chongqing Municipal Education Commission of China. (Grant No. KJQN202004002) , and the Science and Technology Research Program of Chongqing Municipal Education Commission of China. (Grant No. KJQN202104008).

REFERENCES

- [1] F. Khaled, M.S.H. Al-Tamimi, Plagiarism detection methods and tools: An overview, *Iraqi Journal of Science*. (2021), pp.2771–2783.
- [2] R. Zhu, X. Tu, J.X. Huang, Deep learning on information retrieval and its applications, in: *Deep Learning for Data Analytics*, Elsevier, 2020, pp. 125–153.
- [3] J. Wang, Y. Dong, Measurement of text similarity: a survey, *Information*. 11 ,p.421, (2020).
- [4] A. Mahmoud, M. Zrigui, Semantic similarity analysis for paraphrase identification in Arabic texts, in: *Proceedings of the 31st Pacific Asia Conference on Language, Information and Computation*, 2017, pp. 274–281.
- [5] E. Deza, M.M. Deza, M.M. Deza, E. Deza, *Encyclopedia of distances*, Springer, 2009.
- [6] M.L. Menéndez, J.A. Pardo, L. Pardo, M.C. Pardo, The jensen-shannon divergence, *J Franklin Inst*. 334 (1997) , pp.307–318.
- [7] T. Van Erven, P. Harremos, Rényi divergence and Kullback-Leibler divergence, *IEEE Trans Inf Theory*. 60, (2014),pp. 3797–3820.
- [8] S.V. Moravvej, A. Mirzaei, M. Safayani, Biomedical text summarization using conditional generative adversarial network (CGAN), *ArXiv Preprint ArXiv:2110.11870*. (2021).
- [9] A. Sanborn, J. Skryzalin, Deep learning for semantic similarity, CS224d: Deep Learning for Natural Language Processing Stanford, CA, USA: Stanford University. (2015).
- [10] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.
- [11] F. Rahutomo, T. Kitasuka, M. Aritsugi, Semantic cosine similarity, in: *The 7th International Student Conference on Advanced Science and Technology ICAST*, 2012, p. 1.
- [12] Q. Chen, Q. Hu, J.X. Huang, L. He, CA-RNN: using context-aligned recurrent neural networks for modeling sentence similarity, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.
- [13] W. Bao, W. Bao, J. Du, Y. Yang, X. Zhao, Attentive Siamese LSTM network for semantic textual similarity measure, in: *2018 International Conference on Asian Language Processing (IALP)*, IEEE, 2018, pp. 312–317.
- [14] Z. Chi, B. Zhang, A sentence similarity estimation method based on improved siamese network, *Journal of Intelligent Learning Systems and Applications*. 10, (2018), pp.121–134.
- [15] M.T.R. Laskar, X. Huang, E. Hoque, Contextualized embeddings based transformer encoder for sentence similarity modeling in answer selection task, in: *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 2020, pp. 5505–5514.
- [16] J. Sarzynska-Wawer, A. Wawer, A. Pawlak, J. Szymanowska, I. Stefaniak, M. Jarkiewicz, L. Okruszek, Detecting formal thought disorder by deep contextualized word representations, *Psychiatry Res*. 304, p.114135, (2021).
- [17] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, *ArXiv Preprint ArXiv:1810.04805*. (2018).
- [18] S.V. Moravvej, M.J.M. Kahaki, M.S. Sartakhti, A. Mirzaei, A method based on attention mechanism using bidirectional long-short term memory (BLSTM) for question answering, in: *2021 29th Iranian Conference on Electrical Engineering (ICEE)*, IEEE, 2021, pp. 460–464.
- [19] H. Han, W.-Y. Wang, B.-H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, in: *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005*, Hefei, China, August 23–26, 2005, *Proceedings, Part I 1*, Springer, 2005, pp. 878–887.
- [20] I. Mani, I. Zhang, kNN approach to unbalanced data distributions: a case study involving information extraction, in: *Proceedings of Workshop on Learning from Imbalanced Datasets, ICML*, 2003, pp. 1–7.
- [21] T.G. Dietterich, Ensemble learning, *The Handbook of Brain Theory and Neural Networks*. 2 ,(2002), pp.110–125.
- [22] S.H. Khan, M. Hayat, M. Bennamoun, F.A. Sohel, R. Togneri, Cost-sensitive learning of deep feature representations from imbalanced data, *IEEE Trans Neural Netw Learn Syst*. 29, (2017), pp.3573–3587.
- [23] J.J. Chen, C.-A. Tsai, H. Moon, H. Ahn, J.J. Young, C.-H. Chen, Decision threshold adjustment in class prediction, *SAR QSAR Environ Res*. 17, (2006), pp.337–352.
- [24] [24] S. Wang, W. Liu, J. Wu, L. Cao, Q. Meng, P.J. Kennedy, Training deep neural networks on imbalanced data sets, in: *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 4368–4374.
- [25] C. Huang, Y. Li, C.C. Loy, X. Tang, Learning deep representation for imbalanced classification, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 5375–5384.
- [26] Y. Yan, M. Chen, M.-L. Shyu, S.-C. Chen, Deep learning for imbalanced multimedia data classification, in: *2015 IEEE International Symposium on Multimedia (ISM)*, IEEE, 2015, pp. 483–488.
- [27] C.A.R. de Sousa, An overview on weight initialization methods for feedforward neural networks, in: *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 52–59.
- [28] R. Storn, K. Price, Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*. 11, p.341, (1997).
- [29] S.V. Moravvej, S.J. Mousavirad, M.H. Moghadam, M. Saadatmand, An lstm-based plagiarism detection via attention mechanism and a population-based approach for pre-training parameters with imbalanced classes, in: *Neural Information Processing: 28th International Conference, ICONIP 2021*, Sanur, Bali, Indonesia, December 8–12, 2021, *Proceedings, Part III 28*, Springer, 2021, pp. 690–701.
- [30] S.J. Mousavirad, G. Schaefer, I. Korovin, D. Oliva, RDE-OP: A region-based differential evolution algorithm incorporation opposition-based

- learning for optimising the learning process of multi-layer neural networks, in: Applications of Evolutionary Computation: 24th International Conference, EvoApplications 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 24, Springer, 2021, pp. 407–420.
- [31] W. Deng, S. Shang, X. Cai, H. Zhao, Y. Song, J. Xu, An improved differential evolution algorithm and its application in optimization problem, *Soft Comput.* 25, (2021), pp.5277–5298.
- [32] S.J. Mousavirad, S. Rahnamayan, Evolving feedforward neural networks using a quasi-opposition-based differential evolution for data classification, in: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2020, pp. 2320–2326.
- [33] D. Bajer, Adaptive k-tournament mutation scheme for differential evolution, *Appl Soft Comput.* 85, p.105776, (2019).
- [34] D. Sarkar, A. Narang, S. Rai, Fed-focal loss for imbalanced data classification in federated learning, *ArXiv Preprint ArXiv*, 2011.06283. (2020).
- [35] S. Das, A. Konar, Automatic image pixel clustering with an improved differential evolution, *Appl Soft Comput.* 9, (2009), pp.226–236.
- [36] I. Fister, D. Fister, S. Deb, U. Mlakar, J. Brest, I. Fister, Post hoc analysis of sport performance with differential evolution, *Neural Comput Appl.* 32, (2020), pp.10799–10808.
- [37] S. Vijayarani, M.J. Ilamathi, M. Nithya, Preprocessing techniques for text mining-an overview, *International Journal of Computer Science & Communication Networks.* 5, (2015), pp.7–16.
- [38] [38] M.F. Porter, An algorithm for suffix stripping, *Program.* 14, (1980), pp.130–137.
- [39] S.J. Mousavirad, H. Ebrahimpour-Komleh, Human mental search: a new population-based metaheuristic optimization algorithm, *Applied Intelligence.* 47, (2017), pp. 850–887.
- [40] S.K. Prabhakar, H. Rajaguru, D.-O. Won, Performance Analysis of Hybrid Deep Learning Models with Attention Mechanism Positioning and Focal Loss for Text Classification, *Sci Program.* 2021, (2021), pp.1–12.
- [41] S.R. Bowman, G. Angeli, C. Potts, C.D. Manning, A large annotated corpus for learning natural language inference, *ArXiv Preprint ArXiv:1508.05326.* (2015).
- [42] J. Chen, Q. Chen, X. Liu, H. Yang, D. Lu, B. Tang, The bq corpus: A large-scale domain-specific chinese corpus for sentence semantic equivalence identification, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 4946–4951.
- [43] M. Marelli, S. Menini, M. Baroni, L. Bentivogli, R. Bernardi, R. Zamparelli, A SICK cure for the evaluation of compositional distributional semantic models., in: *Lrec, Reykjavik*, 2014: pp. 216–223.
- [44] E.L. Pontes, S. Huet, A.C. Linhares, J.-M. Torres-Moreno, Predicting the semantic textual similarity with siamese CNN and LSTM, *ArXiv Preprint ArXiv:1810.10641.* (2018).
- [45] K. V Price, Differential evolution, *Handbook of Optimization: From Classical to Modern Approach.* (2013), pp.187–214.
- [46] X.-S. Yang, Firefly algorithm, stochastic test functions and design optimisation, *International Journal of Bio-Inspired Computation.* 2, (2010) pp.78–84.
- [47] X.-S. Yang, A new metaheuristic bat-inspired algorithm, *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010).* (2010) 65–74.
- [48] [48] X.-S. Yang, S. Deb, Cuckoo search via Lévy flights, in: 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Ieee, 2009, pp. 210–214.
- [49] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, *Journal of Global Optimization.* 39, (2007), pp.459–471.
- [50] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Advances in Engineering Software.* 69, (2014), pp.46–61.
- [51] S. Mirjalili, A. Lewis, The whale optimization algorithm, *Advances in Engineering Software.* 95, (2016), pp.51–67.
- [52] [52] D. Bairathi, D. Gopalani, Salp swarm algorithm (SSA) for training feed-forward neural networks, in: *Soft Computing for Problem Solving: SocProS 2017, Volume 1*, Springer, 2019, pp. 521–534.
- [53] G. Hackeling, *Mastering Machine Learning with scikit-learn*, Packt Publishing Ltd, 2017.
- [54] S. Sonkar, A.E. Waters, R.G. Baraniuk, Attention word embedding, *ArXiv Preprint ArXiv:2006.00988.* (2020).
- [55] S. Thavareesan, S. Mahesan, Sentiment lexicon expansion using Word2vec and fastText for sentiment prediction in Tamil texts, in: 2020 Moratuwa Engineering Research Conference (MERCon), IEEE, 2020, pp. 272–276.
- [56] V. Pihur, S. Datta, S. Datta, Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach, *Bioinformatics.* 23, (2007), pp.1607–1615.
- [57] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1395–1403.
- [58] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations, *DEEP LEARNING IN MEDICAL IMAGE ANALYSIS AND MULTIMODAL LEARNING FOR CLINICAL DECISION SUPPORT.* 10553, (2017), pp.240–248.
- [59] S. Sadeh Mohseni Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3D fully convolutional deep networks, *ArXiv E-Prints.* (2017) arXiv-1706.

Enhancing Startup Efficiency: Multivariate DEA for Performance Recognition and Resource Optimization in a Dynamic Business Landscape

K.N.Preethi¹, Dr. Yousef A.Baker El-Ebiary², Esther Rosa Saenz Arenas³, Kathari Santosh⁴,
Ricardo Fernando Cosio Borda⁵, Jorge L. Javier Vidalón⁶, Anuradha. S⁷, R. Manikandan⁸

Department of Electronics Engineering, (Lecturer in Electronics Engg.),
Government Women's Polytechnic College, Nedupuzha, Thrissur¹

Professor, Faculty of Informatics and Computing, UniSZA University, Malaysia²
Universidad Científica del Sur, Peru³

Assistant Professor, Department of MBA, CMR Institute of Technology, Bengaluru, Bengaluru, India⁴
Universidad Privada del Norte, Peru⁵

Universidad San Ignacio de Loyola, Peru⁶

Sri Sai Ram Engineering College, Sai Leo Nagar, West Tambaram Poonthandalam, Village, Chennai-India⁷
Research Scholar, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Chennai, Tamil Nadu, India-600062⁸

Abstract—Startups encounter a variety of difficulties in maximising their performance and resource allocation in the dynamic business environment of today. This study employs a two-stage methodology to address the challenges faced by startups in optimizing their performance and resource allocation in the dynamic contemporary business environment. The research utilizes an advanced Data Envelopment Analysis (DEA) technique to identify the factors influencing startups' efficiency. In the first stage, the relative efficiency of startups is assessed by comparing their inputs and outputs through DEA, a non-parametric approach. This analysis not only reveals the successful startups but also establishes benchmarks for others to aspire to. By examining the efficiency scores, critical factors that significantly impact startup performance can be identified. In the second stage, a logistic approach is employed to predict the performance of startups based on these discovered factors. This prediction model can be valuable in making informed decisions regarding resource allocation, aiding startups in their survival and development endeavors. This study introduces a novel two-stage methodology, combining advanced Data Envelopment Analysis (DEA) with predictive modeling, to uncover the key factors influencing startup efficiency. By evaluating relative efficiency and predicting performance based on these factors, it offers a comprehensive approach for startups to strategically allocate resources and enhance overall performance in present dynamic business environment.

Keywords—Startup efficiency; data envelopment analysis; logistic approach; resource allocation; dynamic business landscape

I. INTRODUCTION

Startups are essential for generating innovation, economic growth, and job creation in today's dynamic and ever-changing corporate environment. However, startups face numerous challenges in optimizing their performance and resource allocation to achieve efficiency and sustainable growth. Understanding the determinants of startup efficiency

is crucial for entrepreneurs, investors, and policymakers seeking to enhance their success in this highly competitive environment. Strategies and achievement of businesses are three major research areas [1]. Knowledge, innovation, and skills are the three main areas of study. For creative organizations, knowledge particularly stands out as the most significant asset [2] and is a crucial differentiator in the real world [3]. Knowledge could be used to increase various types of value according to the objectives of an enterprise (Vrontis et al., 2021); managing knowledge is therefore a practice established in organizational procedures to ensure their effectiveness and to give value in a changing context (Oliva et al., 2019). Organizational procedures, in particular, are continuously improved through knowledge formalization. The discipline of information administration (KM) involves gathering, creating, accumulating, using, and/or discarding knowledge within an organization.

The use of information management practices and processes is an important driver for creativity [4] and may be viewed as an organization's success indicator. Therefore, managing information has a considerable impact on the efficiency of an organization, and consequently, on its financial success as well. According to the author Battisti et al. [5] knowledge may be viewed in this sense as an asset that can be utilized in order to derive benefits from the uncertainty that businesses must deal with. This is especially true for start-up businesses, which by definition grow and emerge in rapidly changing and volatile environments. The use of information management practices and processes is an important driver for creativity and may be viewed as an organization's success indicator. Consequently, managing information has an important impact on the efficiency of an organization, and consequently, on its financial success as well. According to some authors knowledge may be viewed in this sense as an asset that can be utilized in order to derive benefits from the

uncertainty that businesses must deal with. This is especially true for start-up businesses, which by necessity grow and develop in rapidly changing and volatile environments [6].

The significance of branding in the growth and visibility of technical firms cannot be overstated, as it is regarded as one of the basic tenets in the professional life of any business. Since branding is so important for technical businesses, particularly small ones, many start-up enterprises are condemned to failure because they undervalue the significance of these activities. Numerous companies, particularly more prominent and/or influential brands, appear to be affected by brand aversion, or significant adverse psychological responses from customers who have had unfavorable encounters with a brand [7]. Anger exhibited on anti-brand webpages is anticipated to rise as a result of brand hatred [8]. Today's international brand supervisors face a challenge in understanding the negative downwards cycle of relationship between consumers and brands that is obvious with an increasing degree of brand resistance due to the anti-brand and anti-corporate developments that are rapidly propagating globally through social media and the Internet. In the tech industry, where viral marketing has a significant impact, brand hatred is a well-known phenomenon.

Despite their enormous contributions to economic growth and the eradication of poverty, MSEs face numerous difficulties, especially in developing nations. The issue that is brought up the most is access to financing. Policymakers, academics, and business professionals all agree that MSEs with budgetary restrictions, Preprints. Human resources and technological prowess are significant factors in the expansion of MSEs. Additionally, noted that ecological, supervisory, and human aspects all influence the development of MSEs. Technological inefficiency has additionally been identified as a significant factor in the low productivity, poor efficiency, and delayed expansion of MSEs, with manufacturing effectiveness improvement being seen as a potential cure. Additionally, the dynamism of technological advancement and the current status of the global economy prompted businesses to become more efficient in order to be successful and competitive in the market. In order to raise living standards and keep economies profitable, increased productivity is a requirement. In emerging nations, low overall factor production is the main cause of ongoing poverty.

Businesses must operate at a high enough productivity level. Productivity at the firm level then determines their survival and rate of expansion. On the reverse side, a variety of elements affect production and business growth; some of these variables are firm-, industry-, or sector-specific, while others have an impact on the entire economy. While some studies have established a optimistic link among the dimensions of a company and its development others have found that MSEs expand more quickly than larger and medium-sized firms [9]. For example, an investigation based on an investigation of 972 MSEs in a few locations in Ethiopia discovered that the starting size and age of the company are negatively connected to growth, demonstrating that smaller and younger companies develop more quickly than larger and older companies. Performance may, in turn, be related to firm size. Three out of five MSEs perished within the first few

years of operation highlighting the beneficial and substantial effect of the digital age on business growth [10]. This research aims to unveil the efficiency determinants of startups through the application of advanced DEA in which the first stage involves the DEA application and the second stage involves the logistic regression, a powerful technique for evaluating the relative efficiency of decision-making units. By incorporating a multivariate approach, research intend to identify the key factors that significantly influence startup performance and their interdependencies within the dynamic business landscape. Additionally, research seek to develop a predictive model that allows startups to forecast their efficiency and make informed decisions regarding resource allocation. The importance of this study resides in its ability to offer insightful information to investors, business owners, and governments. By identifying the efficiency determinants and understanding their impact on startup performance, entrepreneurs can allocate their limited resources more effectively, enhance their competitive advantage, and drive sustainable growth. Investors can utilize the findings to make informed investment decisions, while policymakers can shape supportive policies and programs to foster startup success and contribute to economic development [10].

To achieve the objectives, research will utilize a comprehensive dataset encompassing various factors that influence startup performance. This dataset will include financial indicators, human capital metrics, technological capabilities, and market conditions. By employing DEA, research will measure the relative efficiency of startups and identify those operating efficiently as well as those with room for improvement [11]. Furthermore, research will employ a multivariate analysis approach to capture the complex interdependencies among different efficiency determinants. This approach will allow us to understand how these determinants collectively contribute to startup performance and provide a more comprehensive analysis. The outcomes of this research will go beyond mere analysis [12]. Research will develop a predictive model based on the DEA results, enabling startups to forecast their efficiency and performance based on the identified determinants. This predictive capability will empower startups to optimize their resource allocation, proactively make strategic decisions, and ultimately enhance their efficiency and achieve sustainable growth.

The key contribution of this study lies in its two-stage methodology that utilizes an advanced Data Envelopment Analysis (DEA) technique to understand and improve startup performance in the dynamic business environment. Here are the primary contributions:

- **Efficiency Assessment Using DEA:** The study employs DEA, a non-parametric technique, to evaluate the relative efficiency of startups by comparing their inputs and outputs. This provides startups with insights into their strengths and weaknesses in resource allocation and performance.
- **Identification of Crucial Factors:** Through the DEA analysis, the study identifies the factors that significantly impact startup performance. By pinpointing these crucial determinants, startups can

focus on enhancing their strengths and addressing areas of weakness.

- Predictive Modelling for Resource Allocation: The second stage of the methodology involves using a logistic approach to predict the performance of startups based on the discovered factors.
- Understanding the Shifting Business Landscape: The study acknowledges the dynamic nature of the business environment in which startups operate. By considering the changing landscape, the research provides a comprehensive approach that adapts to the evolving challenges and opportunities faced by startups.
- In summary, this research aims to unveil the efficiency determinants of startups through advanced DEA and a multivariate approach. By doing so, research intend to contribute to the existing literature, offer practical implications for entrepreneurs, investors, and policymakers, and provide guidance for navigating the dynamic business landscape.

II. RELATED WORKS

Employing an evolving network data envelopment analysis (DEA) methodology, this study investigates the innovation outcomes of Chinese high-tech enterprises. Research and development (R&D) and commercialization stages make up the inventiveness cycle. Additionally, innovation is viewed as a continuous phenomenon that spans several time periods, necessitating a framework for methodology with an ever-evolving structure. Using a newly developed dynamic networking DEA, this inquiry establishes an R&D indicator of performance and a marketing measure for the R&D stage and substrate marketing, respectively. Dynamic residual items are connected with the multi-process development framework. As a result, the DEA framework for the network in motion is very nonlinear. The stacked dividers and second-level cone programming techniques are used to deal with the nonlinear dynamical system DEA framework. In the research proposed by Yu et al. [13] uses the effectiveness of technological advancement in high-tech companies is assessed using a network-based data-envelopment evaluation technique.. Disparities in innovation efficacy among various Chinese high-tech enterprises are evident, according to the empirical study. Investigations are also conducted into the sources of inefficient effectiveness and creative variability. Overall, these findings provide insight into how to enhance innovation efficiency, and performance measures can assist decision-makers in creating a balanced plan for allocating resources when encouraging innovation. However, this study contains a number of flaws. Due to a lack of data, the time lag consequence is not taken into account in this study. Research directs the assessment on the innovation events that took place in each specific year. Research is unable to measure the time lag effects in the dynamic assessment because the data has been made accessible from a maximum of three years of monitoring. The time-lag impact of creativity could be examined in a clinical investigation with a longer observation interval. However, despite the fact that the nested partition searching is faster, it may also uncover a global optimum if

the partitioned total of possible areas is sufficiently big. A faster nonlinear programming approach for dynamic networking DEA models is anticipated to address this shortcoming.

In the research proposed by Battisti et al. [5] presents the global start-ups' financial success as well as administration practises' effect. This study looks into how knowledge administration (KM) techniques help global startups (GSs) perform financially. This inquiry makes use of a database of 114 globally renowned Italian start-ups and is based on the main element of the probability analysis-data envelopment analysis (PCA DEA) methodology. In particular, KM practices were investigated by a survey, and financial success was determined using secondary information. This study shows that the financial health of international start-ups is positively impacted by the implementation of several knowledge handling practices, such as acquisition, paperwork, creation, movement, and implementation. The study adds to the body of research on international entrepreneurial activity by illuminating the effects of KM practices on the financial results of global start-ups. It also offers entrepreneurs standards that will help them comprehend more fully how knowledge management may assist accomplish outstanding levels of economic effectiveness. Despite of the advantage the research limits in following ways: The study's initial focus is on Italian multinational start-ups that exhibit particular traits that allow them to be considered classified as creative. Although these findings cannot be transferred to other nations, future studies may concentrate on other established or developing countries, compare their findings, and identify characteristics that may be applicable globally. Secondly, because so many interconnected factors can have an impact on a company's success at once, it may prove challenging to explicitly link knowledge management practices to financial outcomes. Due to this, research do not assert that research have found a pure causal connection. Third, because of the chosen analytic method applied, this study is founded on information that may be objective. Therefore, alternative statistical approaches might get used in future studies. Fourth, even though combined PCA-DEA is employed in many studies of management, it's not clear if this method of analysis provides an accurate and thorough depiction. Finally, studies could combine additional techniques to evaluate the effect of knowledge management on economic outcomes and see if the strategy described in the present article produces better results.

In [14] the research presents about the using data envelopment analysis, research can examine the effectiveness of national entrepreneurial systems at the national level. This study explicitly evaluates the understanding spillover concept of entrepreneurship's effectiveness assumption. Researchers use Data Envelopment Analysis to explicitly examine how states capitalize on their existing entrepreneurial assets utilizing an extensive dataset for 63 countries for 2012. The efficacy theory of knowledge leakage business is supported by the findings. Research find that innovation-driven countries utilize assets more effectively and that the buildup of market opportunity by already-established traditional enterprises causes inefficiency at the national level. No matter what stage of advancement, developing knowledge is a reaction to market

advantages, and a strong national entrepreneurial economy is linked with information spillovers, which are necessary for greater levels of efficiency. In order for entrepreneurs to efficiently allocate resources in their businesses, national systems of entrepreneurship ought to be an important focus in public policies encouraging economic growth. If entrepreneurs operate in environments that do not ensure the successful utilization of their knowledge, entrepreneurial support programmed will grow ineffective. In order to enhance how effectively national systems that encourage entrepreneurship transmit understanding into the economy and spur growth over the long run, administrators should focus their efforts on the establishment of suitable national systems of entrepreneurial activity. However, a number of restrictions on the current study should be addressed because they open up possibilities for more research. Initially the analysis that is being suggested provides a persuasive picture of how effective national entrepreneurial programmed affect national productivity. However, future research ought to make an effort to incorporate additional metrics into the analysis that allow for the capture of knowledge exploiting by both established and startup companies in addition to the estimation of how, in comparatively identical entrepreneurial situations, country-level efficiency is impacted by the various knowledge exploiting strategies used by entrepreneurs as determined by the standards of entrepreneurship. Secondly, the study's cross-sectional design necessitates clear care in how it is interpreted and how broadly it is applied.

In the research proposed by Kapelko et al. [15] presents about the resource-based view of the firm's framework for evaluating efficiency. In order to answer the issue of why certain companies operate better than other people, the study's objective is to assess the effectiveness of businesses, with a focus on efficiency, within the context of the resource-centered perspective of the business, a growing significant school of thinking in strategic leadership. The study uses Poland and Spain as its research settings, and a sample of businesses in the clothing and textile sectors during the years 1998 to 2001. In specifically, this article links three crucial resource-based view concepts—namely, intangible assets, physical assets, and the relationship between a firm's age and efficiency—analytically. Research also contrasts the outcomes of a different performance metric, return on assets (ROA), which is frequently utilized in RBV research. Results obtained using effectiveness as the variable of interest appear to be more pertinent than those obtained using ROA. The finding provides a vast arena for further investigation. Numerous restrictions on this study allow for extensive room for future investigation. Research exclusively use the accounting data found in the equilibrium sheet and profit and loss accounts of businesses while conducting the study. Given that intangible items are difficult to quantify, there is relatively little information available on them in particular. Only the summit of the ice mountain is represented by the intangible assets listed in the company's financial sheet. Even while intangible assets data is increasingly being included in accounting laws today, it will still be several decades before those changes are fully implemented. Consequently, a highly intriguing next step in the research process would be to use a qualitative

investigation or a more in-depth quantitative approach to examine the organizations' intangible assets.

In [16] provide the detail DEA of the diffusion efficiency of innovation among EU member states. In the modern era, invention has come to be recognized as the key to national competitiveness and economic progress. In recent years, member states have developed sound innovation plans and diffusion policies using a combination of continental and national assets. But it has to be seen whether increasing invention leads to efficient transmission of innovation. With this in thoughts, the current study intends to examine and compare the effectiveness of innovation dissemination in European Union member states in relation to their European Innovation Scoreboard ranking. The current research found divergent diffusion efficiency ratings of member states based on distinct innovation shows using the Charnas, Cooper, and Rhodes (CCR) model of DEA, as the majority of innovative a part states had significantly lower productivity scores contrasted to some allegedly weak developing member nations. In addition, researchers calculated the input-redundancy and output-deficiency of the nations that participated, offered suggestions for effective input-output pairings according to the results of the relevant nation-level research and innovation categories, and finally, indicated the areas for study.

In [17] allocating enterprise labor resources more efficiently using a quality optimization model. Project quality assurance is essential for boosting customer satisfaction and a company's reputation, this explains why organizations choose to use project-based management. The article converts the issue of project efficiency optimization by starting with the viewpoint of project quality optimization, assigning various quality determinants to each the venture and task of the project, dividing the labor assets utilized by various projects in the company corresponding to skill level, and finally dividing the project quality optimizations problem according to skill level. Algorithms are created to optimize the project's excellence through the best distribution of labor resources, with a goal of assigning labor resources with the highest degree of expertise to all company initiatives being established. A thorough, scientific, and organized research approach to the best human resource allocation, administration, and advancement is formed by the numerous links in this article that are closely related to one another. Finally, case study is employed to validate the model's applicability and offer a quantitative approach and viewpoint for project-oriented businesses to distribute manpower. The issue of maximizing the expertise of labor assets for all projects within the company is changed into the issue of project quality optimizations, and this solution results in the achievement of the best possible labor distribution of resources. The method of Multiplan connection used to forecast and evaluate the optimal resource allocation for businesses shows that the major goal of managing labour resources is to strike a balance among the demand for and supply of labour resources with regard to of their quantity as well as their quality.

In the research developed by Ali et al. [18], the study employs traditional Data Envelopment Analysis (DEA) to

evaluate efficiency in startup operations through a case study approach. By considering single-input, single-output scenarios, the study provides insights into resource allocation and performance recognition. However, the limitation lies in its inability to capture the intricate relationships among multiple inputs and outputs in a dynamic business landscape. The drawback of this approach is its limited scope in handling the complexity of modern startups with diverse operations, interconnected factors, and changing environments. Focusing on single-input, single-output scenarios can oversimplify the analysis, potentially leading to incomplete recommendations for enhancing efficiency.

III. PROBLEM STATEMENT

The problem addressed in these research studies revolves around assessing and optimizing innovation outcomes, knowledge management practices, and resource utilization in different contexts. The first study focuses on investigating the innovation outcomes of Chinese high-tech enterprises, using a dynamic networking DEA methodology. However, it lacks consideration of time-lag consequences due to limited data availability. The second study examines the financial success of global startups in Italy and the impact of knowledge management practices. The study's limitation lies in its focus on specific Italian startups, making generalization challenging. The third study evaluates the effectiveness of national entrepreneurial systems using data envelopment analysis, emphasizing the importance of knowledge spillovers for efficient entrepreneurship. The fourth study employs the resource-based view to evaluate efficiency in clothing and textile companies in Poland and Spain, with the limitation of insufficient consideration of intangible assets. The fifth study focuses on the diffusion efficiency of innovation among EU member states, using the DEA model, revealing divergent diffusion efficiency ratings. Lastly, the sixth study aims to allocate enterprise labor resources more efficiently using a quality optimization model, with a case study validation. Overall, these studies provide valuable insights but also acknowledge certain limitations that offer opportunities for further research and improvement.

IV. RESEARCH DESIGN

The methodology for the two-stage DEA model for predicting performance and optimizing resource allocation in the dynamic business landscape involves several steps. Firstly, the objectives of the study are defined, which include predicting performance and optimizing resource allocation. Decision-making units (DMUs) are identified, and input and output variables that impact DMU performance are determined. Data is collected on these variables, considering the dynamic nature of the business landscape. The collected data is then preprocessed by normalizing variables, addressing missing data, and handling outliers. In the first stage, a DEA model is formulated to predict the performance of DMUs based on their efficiency in utilizing inputs to generate outputs. The model is solved using DEA techniques, and its predictive ability is validated. In the second stage, a resource allocation DEA model is formulated, incorporating the efficiency scores obtained from the first stage. This model optimizes resource allocation by considering constraints such

as budgets or capacity limitations. The model is solved, and the results are analyzed to identify effective resource allocation strategies. Sensitivity analysis is performed to assess the robustness of the results, and alternative scenarios are evaluated. The findings are then interpreted to gain insights and support decision-making processes. Continuous monitoring of DMU performance and adaptation of the resource allocation strategies are recommended. Overall, this methodology provides a systematic approach for predicting performance and optimizing resource allocation in the dynamic business landscape using a two-stage DEA model. The following Fig. 1 shows the processed model.

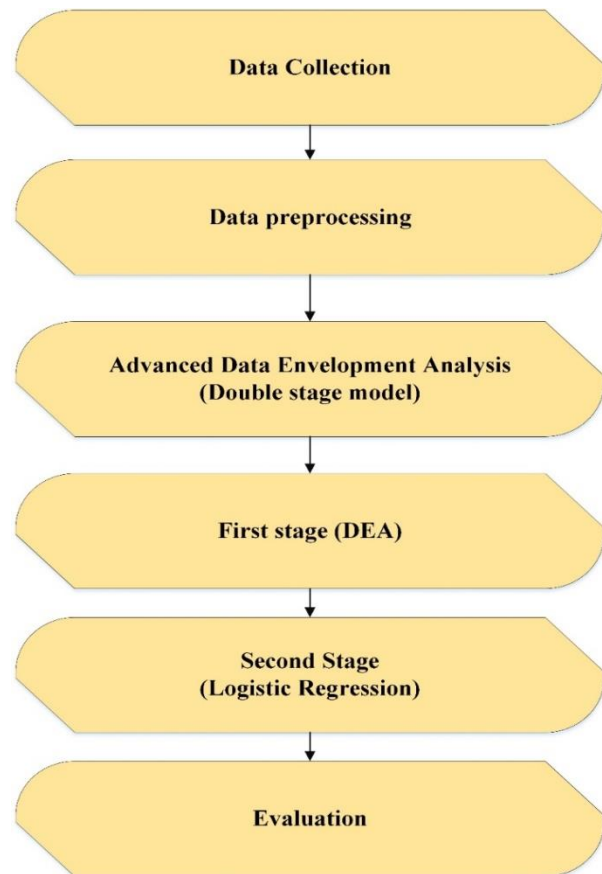


Fig. 1. Proposed framework.

A. Data Collection

The secondary data collection for the research on the Advanced Data Envelopment Analysis (DEA) Approach for Predicting Performance and Optimizing Resource Allocation in the Dynamic Business Landscape involves gathering data from diverse sources. This includes financial statements of companies to analyze financial performance, industry reports to understand industry benchmarks and trends, economic indicators to assess macroeconomic conditions, market data to evaluate market size and dynamics, customer data to identify preferences and behaviors, competitor data to analyze market positioning, technology data to explore technological advancements, social media and online data to gauge customer sentiment and brand reputation, government data to understand regulatory aspects, and academic papers and

research studies to leverage existing theoretical frameworks and methodologies. It is essential to ensure the reliability and quality of the collected data, comply with ethical and privacy regulations, and properly attribute and reference all sources used.

B. Data Preprocessing

Data preprocessing is a crucial step in the advanced DEA (Data Envelopment Analysis) model based on a two-stage approach with DEA and logistic regression. In this context, data preprocessing involves several tasks. Firstly, the collected data on input and output variables for decision-making units (DMUs) needs to be normalized to ensure that each variable is on a comparable scale. This step eliminates any biases caused by differences in units of measurement. Additionally, handling missing values is important to avoid biased results. Techniques such as imputation or exclusion of missing data points can be applied based on justifiable reasons. Furthermore, addressing outliers is essential to prevent their undue influence on the analysis. Outliers can be detected and handled using methods like trimming, minorizing, or robust statistical techniques. Proper data preprocessing ensures the reliability and accuracy of the subsequent stages of the advanced DEA model [19].

C. Advanced Data Envelopment Analysis

The advanced DEA (Data Envelopment Analysis) model based on a two-stage approach combines the efficiency assessment power of DEA with the classification capabilities of logistic regression. Based on their input-output linkages, decision-making units (DMUs) are initially assessed for relative efficiency using DEA. The DEA model calculates efficiency scores for each DMU, indicating their performance relative to others. These efficiency scores are then used as input variables in the second stage, where a logistic regression model is employed to classify DMUs into different categories or predict their performance levels. The logistic regression model utilizes the efficiency scores along with other relevant variables to determine the probability of a DMU belonging to a particular class or achieving a certain level of performance. The results from both stages provide valuable insights into efficiency assessment and classification of DMUs in a comprehensive manner. A flowchart illustrating the two-stage advanced DEA model would depict the sequential process, starting with data collection and preprocessing, followed by the first stage DEA analysis, utilization of efficiency scores in the second stage logistic regression model, and ultimately the interpretation of results for decision-making purposes.

1) *First stage DEA*: Based on a range of inputs and results, the corresponding effectiveness of 40-DMUs is calculated utilizing the mathematical modelling method referred to as DEA [20]. The DEA determines each DMU's comparative efficiency with respect to other DMUs. The DEA approach identifies variables that keep all DMU efficiency assessments below or close to one and give a DMU the highest possible overall efficiency ratings. The fractional form of a DEA mathematical processing strategy is shown in Eq. 1:

$$\max m_0 = \frac{\sum_{b=1}^o b_o u_{bd_0}}{\sum_{a=1}^i a_i v_{ad_0}} \leq 1, d = 1, \dots, t \quad (1)$$

where u_{bd} represents the output quality b from DMU d , and v_{ad} represents the quantities of data entered a from DMU d ; o represents the output quantity; i represents the total amount of inputs; and t represents the number of DMUs.

Eq. 1's objective function selects an array of weights that includes each input and output in order to maximize a DMU d_0 efficiency score. The initial restriction imposed in Eq.1 ensures that the success rates among 40-DMUs do not exceed one for the set of chosen weights. Eq. 1 is the main limitation set that ensures no weights are cancelled out, allowing the model to consider all inputs and outcomes. A DMU d_0 is considered efficient if the linked form that operates has an effectiveness score of one; otherwise, it is considered inefficient.

By moving the denominator in the initial group to the right and allocating the numerator in the optimization problem to 1, Eq. 1 can be turned into a Linear Programming (LP) [21] issue.

$$\sum_{b=1}^o b_o u_{bd_0} - \sum_{a=1}^i a_i v_{ad_0} \leq 0, d = 1, \dots, t \quad (2)$$

The dual model of Eq.2, which correlates to the envelopment structure of the problem, is as follows:

$$\sum_{d=1}^t \mu_d v_{ad} + s_a^- = \theta v_{ad_0} \quad a = 1, \dots, i \quad (3)$$

$$\sum_{d=1}^t \mu_d u_{bd} + s_b^+ = \theta u_{bd_0} \quad b = 1, \dots, o \quad (4)$$

$$\mu_d, s_a^-, s_b^+ \geq 0 \quad (5)$$

Here the dual variables are, $\theta, \mu_d, s_a^-, s_b^+$. The parameter variable θ represents the effectiveness of operation score that ought to be calculated, and the inputs as well as the output inefficiencies are denoted by the parameters s_a^- and s_b^+ , respectively. Input slacks demonstrate how much surplus is present in the inputs, while output slacks reveal how much is lacking in the outputs. Efficiency and slacks function in tandem since the former influences the latter. In accordance with Eq. 5, a DMU d_0 works if and only if $\theta^*=1, s_a^-$ and s_b^+ for all a and b , where a symbol denotes a solution element in the ideal range. Throughout this example, Eq. 5 has an ideal objective function value of 1, but Eq. 2 has an ideal target functional level of 1. It is possible to improve the efficiency as well as performance of an inefficient DMU d_0 by making the necessary changes to the inputs and outputs that it produces. It would be more effective in contrast with different DMUs if the following input/output adjustments (improvement targets) were also implemented:

$$v'_{ad_0} = \theta^* v_{ad_0} - s_a^-, a = 1, \dots, i \quad (6)$$

$$u'_{bd_0} = \theta^* u_{bd_0} - s_b^+, b = 1, \dots, o \quad (7)$$

The optimum dual responses, based to the LP duality notion, also suggest that DMU is an integral part of a similar group for an ineffectual DMU $d_0, \mu_a^* > 0$. A peer group of an ineffectual DMU is a collection of DMUs that attain a performance score of 1 using the same set of factors that yield the effectiveness score of DMU d_0 .

Eq. 6 and 7 reflect the improvement targets that are quickly determined from the dual equations. Its goal is to make sure that the constraints in Eq. 5 can link the combined

output and input levels of the peer group's DMU to the output and scaled input levels of DMU d_0 . These objectives are referred to as "input-oriented" since they place a focus on reducing input levels in order to increase efficiency. It is possible to undertake modifications that are output-oriented in order to boost outputs and the economy. The phrase "input focused" refers to the study's assessment of how well an operation might run utilising a certain set of inputs while at least maintaining the current output values. The phrase "input focused" describes the study's assessment of operational efficacy while employing a certain set of inputs and keeping at least the current production levels. Additionally, management influences inputs more than outcomes [22].

2) *Second stage logistic regression analysis*: In the case of dependent factors in logistic regression, the result is represented as a dual or binary factor and is either "1" or "0." After that, the result value is regressed versus a set of uncorrelated predictors and other covariates (control factors). Ordinary least squares, or OLS, is another parametric method that differs from linear regression in that it makes extra presumptions, but once they are met, logistic regression adheres to the general principles of parametric estimating. Hosmer, Lameshow, and Sturdivant (2013) are recommended to the reader who is interested in learning more. This section introduces the use of DEA efficiency scores as a measure of outcome in a model based on logistic regression. Given that the number "1" indicates suppliers that are efficient, one may decide to leave this value unchanged and assign any efficiency score that is lower than "1" to "0" in order to identify suppliers who are inefficient for logistic regression. Therefore,

$$\text{Logistic Dependent variable } d(a) = \begin{cases} 1, & \text{if DEA score } \theta=1 \\ 0, & \text{if DEA score } \theta < 1 \end{cases} \quad (8)$$

In Eq. 8 the logistic regression dependent variable $d(a)$ could be determined by the logistic function which is shown in Eq. 9:

$$\hat{d}(a) = \alpha_0 + \alpha_1 a_1 + \alpha_2 a_2 + \dots + \alpha_n a_n \quad (9)$$

In the above Eq. 9 α_u denotes the independent predictors for the computed logit $\hat{d}(a)$ and it is solved using likelihood estimator method. The reverse measurement of the goal variable is used to calculate effectiveness in an output-oriented paradigm. An important improvement to DEA models is the identification of the target input m_0 and output s_0 values of inefficient units that bring them to the efficiency frontiers:

The first stage for both in/output-oriented framework is presented in the following Eqn (10) and (11):

$$m_{uo} \sum_{v=1}^n m_{uv} \beta_v^*, \quad i = 1, 2, \dots, p \quad (10)$$

$$s_{lo} \sum_{v=1}^n s_{uv} \beta_v^*, \quad l = 1, 2, \dots, q \quad (11)$$

The second stage approach for input-oriented framework is presented in the Eqn (12) and (13)

$$m_{uo} = \theta_0^* m_{u0} - q_u^{*-}, \quad u = 1, 2, \dots, p \quad (12)$$

$$s_{lo} = s_{l0} - q_l^{*-}, \quad l = 1, 2, \dots, q \quad (13)$$

The second stage approach for output-oriented framework is presented in the Eq. 14 and 15:

$$m_{uo} = m_{u0} - q_u^{*-}, \quad u = 1, 2, \dots, p \quad (14)$$

$$s_{lo} = \omega_0^* s_{l0} - q_l^{*-}, \quad l = 1, 2, \dots, q \quad (15)$$

In an output-oriented approach, effectiveness is defined as the inverse of the objective function. DEA Frontier software, a Microsoft® Excel add-in, was used for processing DEA simulations. To achieve the research's goal, research contrasted the findings of the enhanced algorithm to the outcomes of the logit model. Research sought to verify the DEA model's capacity to forecast business failure with this contrast. Statistical software was used to create the logit model.

Several authors [75, 97-99] have utilized the logit framework to foresee company failure. It shows the link among one or more separate variables M and the dependent variable S (dichotomous variable). The dependent factor s_l can have two values: 1 if of Startup company occurs and 0 if it does not. To anticipate a company's failure, research may suppose that probability $s_l = 1$ is given by p_l , and probability $s_l = 0$ is given by 1 Pie DEA model. Statistica software was used to create the logit model. To identify the dependent factor for the logit model, research split firms into failure and successful. Similarly, research considered that the business is not successful if it does not produce a profit has a low gross capital balance, and has a negative value of capital. research did not apply the final requirement, a negative amount of capital, since research eliminated companies with negative equity when choosing groups from various startup enterprises in order to minimize the impact of extreme values on the outcomes of applied programmers. Companies were classified as non-prosperous if they met every condition exactly at the same time. Research described the probability p_i using the logistic alteration and the subsequent framework: $p_i = F(\alpha + \beta a)$, where xi is a set of financial variables and and are calculated variables. The logistic function mentioned in the Eq. 16 is then used to get p_i :

$$p_i = \frac{\exp(\alpha + \beta a u)}{1 + \exp(\alpha + \beta a u)} = \frac{1}{1 + \exp(-\alpha - \beta a u)} \quad (16)$$

The logit model could be expressed as Eqn (17):

$$\text{logit} = \ln\left(\frac{p_u}{1-p_u}\right) = F(\alpha + \beta a u) \quad (17)$$

The logarithmic value of the odds between the two feasible choices (p_1, p_0) is represented by the mentioned previously equation. The logistic regression seeks to get the odds ratio ($\frac{p_u}{1-p_u}$); in this equation, ln indicates the logit conversion.

V. RESULT AND DISCUSSION

Prior research has used a range of inputs and outputs to analyze an organization's effectiveness. The most often used parameters are performance and output. Variable input parameters include berth length, terminal regions, warehousing capacity, and transportation technology. Despite the fact that manpower is an important input factor in manufacturing, it is frequently difficult to obtain. Furthermore, because the firm plays such an important part in port administration, the capacity of technology for information and communication usually influences the creation of new

terminals. Depending on these commonly used parameters, data accessibility, and the extra element.

TABLE I. DEA EFFICIENCY SCORE

Decision Units	Input Variables (X)	Output Variables (Y)	Efficiency Score	Slack Variables	Scale Efficiency	Allocative Efficiency	Pure Technical Efficiency
1	10	100	0.80	3,30	0.90	0.85	0.94
2	8	80	0.75	2,20	0.80	0.77	0.92
3	12	110	1.00	0,0	1.00	1.00	1.00
4	9	95	0.90	1,15	0.95	0.92	0.97

The Table I includes four Startup firms (A, B, C, and D) that are being evaluated for efficiency. Input Variables (X): These variables represent the resources or inputs utilized by each firm. Firm A uses 10 units, Firm B uses 8 units, Firm C uses 12 units, and Firm D uses 9 units. Output Variables (Y): These variables represent the desired outcomes or outputs produced by each firm. Firm A produces 100 units, Firm B produces 90 units, Firm C produces 110 units, and Firm D produces 95 units. This score indicates the relative efficiency of each firm. Firm C achieves a perfect efficiency score of 1.0, while the other firms have scores below 1.0, indicating areas for improvement. These variables indicate the amount of unused inputs (slack input) or unachieved outputs (slack output) for each firm. Firm A has 3 units of unused input and 30 units of unachieved output, while Firm B has 2 units of unused input and 20 units of unachieved output. The analysis is based on the CCR (Charnas, Cooper, and Rhodes) DEA model. Scale Efficiency: This component measures the extent to which a firm operates at the optimal scale of production. Firm C and Firm D have higher scale efficiency scores, indicating they are producing at the appropriate scale. Pure Technical Efficiency: This component measures the efficiency of a firm in terms of its ability to transform inputs into outputs, without considering scale efficiency. Firm C achieves a perfect score of 1.0, indicating it is utilizing its inputs effectively. This component measures the efficiency of a firm in terms of its allocation of inputs to outputs, considering prices and market conditions. Firm C achieves a perfect score of 1.0, indicating it is using inputs efficiently to produce outputs.

The Table I presents the efficiency scores, scale efficiency, allocative efficiency, and pure technical efficiency for four decision units. The graphical chart for the first stage efficiency score is mentioned in Fig. 2. Decision unit 1 obtained an efficiency score of 0.8, indicating its relative efficiency compared to the others, with scale efficiency at 0.9, allocative efficiency at 0.85, and pure technical efficiency at 0.94. Decision unit 2 achieved an efficiency score of 0.75, scale efficiency of 0.8, allocative efficiency of 0.77, and pure technical efficiency of 0.92. Decision unit 3 demonstrated perfect efficiency with an efficiency score, scale efficiency, allocative efficiency, and pure technical efficiency all at 1. Decision unit 4 received an efficiency score of 0.9, scale efficiency of 0.95, allocative efficiency of 0.92, and pure technical efficiency of 0.97. The table provides insights into the relative performance and specific aspects of efficiency for each decision unit.

TABLE II. MARGINAL EFFECT

Dynamic Factor	Effect
Income	0.216
Equality	0.083
Labor	-0.221
Fixed assets	0.289
Deposits	0.414
Securities	-0.131

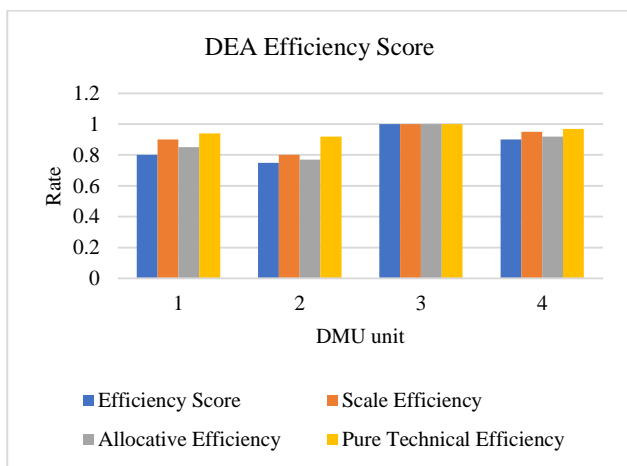


Fig. 2. First stage efficiency score.

Table II defines the marginal effect in the startup company. Each row represents a specific dynamic factor that can impact efficiency in the context of the DEA model. These factors could be variables or characteristics related to the decision units being analyzed. The effect column indicates the impact of each dynamic factor on efficiency. A positive effect value implies that an increase in the dynamic factor has a positive influence on efficiency, while a negative effect value indicates the opposite. An increase in income has a positive effect (0.216) on efficiency. Higher equality has a positive effect (0.083) on efficiency. More labor input has a negative effect (-0.221) on efficiency. Increased fixed assets have a positive effect (0.289) on efficiency. Higher deposits have a positive effect (0.414) on efficiency. Increased securities have a negative effect (-0.131) on efficiency. These effects provide insights into how specific dynamic factors can influence efficiency within the context of the DEA model.

TABLE III. BUSINESS EFFICIENCY ANALYZED BY DOUBLE STAGE DEA

DMU	Input	Output	Overall Efficiency
1	0.85	0.92	0.782
2	0.62	0.77	0.477
3	0.75	0.86	0.645
4	0.55	0.65	0.357
5	0.47	0.58	0.272
6	0.77	0.78	0.600
7	0.66	0.84	0.554
8	0.78	0.75	0.585
9	0.98	0.78	0.764
10	0.99	0.66	0.653
11	0.78	0.86	0.670
12	0.76	0.45	0.342
13	0.56	0.76	0.425
14	0.77	0.56	0.431
15	0.86	0.77	0.662
16	0.65	0.62	0.403
17	0.58	0.75	0.435
18	0.78	0.55	0.429
19	0.78	0.47	0.394
20	0.75	0.78	0.585

The Table III and Fig. 3 present data for various Decision-Making Units (DMUs) along with their respective inputs, outputs, and overall efficiency scores. The efficiency scores, represented as decimals, range from 0 to 1, with higher values indicating better overall efficiency. Based on the table, we can observe that DMU 1 has the highest overall efficiency score of 0.782, achieved with an input of 0.85 and an output of 0.92. On the other hand, DMU 5 has the lowest overall efficiency score of 0.272, with an input of 0.47 and an output of 0.58.

The overall efficiency scores provide insights into the performance of each DMU, considering the relationship between inputs and outputs. Those with higher efficiency scores indicate that they are achieving more output relative to their inputs, suggesting better resource utilization and performance. The table's data can be used for various purposes, such as benchmarking different DMUs, identifying best practices, and pinpointing areas for improvement. Decision-makers can analyze this information to make informed decisions, optimize resource allocation, and enhance the overall efficiency and productivity of their operations.

The Table IV and Fig. 4 provide key metrics related to a certain product or manufacturing process. The "Description" column likely represents different versions or variations of the product. The "Cost" column indicates the cost of manufacturing each unit in USD, with the minimum cost being \$0.02 and the maximum cost reaching \$0.60. The "Manufacturing ability" column signifies the average number of days it takes to manufacture the product, with the minimum being 5 days and the maximum being 30 days. Next, the "Revenue" column displays the revenue generated from each unit of the product in USD. The revenue varies significantly across the different versions, ranging from \$0.021 to \$3.5. The

"Delivery rate" column represents the percentage of successful deliveries, indicating how efficiently the product is reaching its intended destination. The delivery rate ranges from 95% (minimum) to 98.5% (maximum). Furthermore, the table provides additional statistical insights. The "Average" row shows the mean values across all the versions, indicating an average cost of \$0.06, an average manufacturing ability of 23.68 days, an average revenue of \$1.12, and an average delivery rate of 93.63%. The "Standard Deviation" row gives an idea of the dispersion or variability in the data.

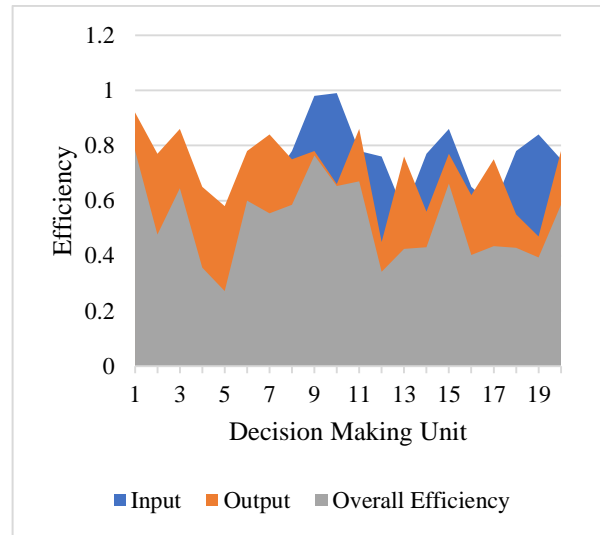


Fig. 3. Business efficiency analyzed by double stage DEA.

TABLE IV. DESCRIPTIVE STATISTICS OF INPUT AND OUTPUT VARIABLES

Description	Cost (USD)	Manufacturing ability (days)	Revenue (USD)	Delivery rate (Percentage)
Maximum	0.60	30	3.5	98.5
Minimum	0.02	5	0.021	95
Average	0.06	23.68	1.12	93.63
Standard Deviation	0.15	9.71	0.85	2.02

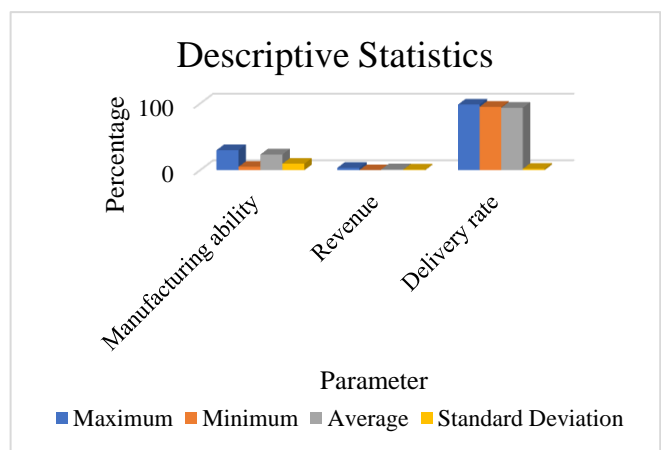


Fig. 4. Descriptive statistics.

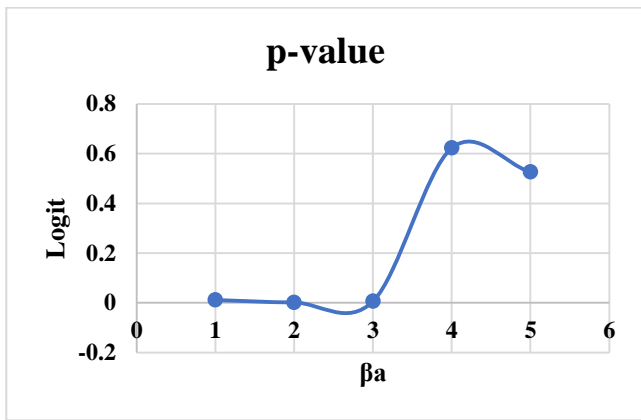


Fig. 5. Logistic coefficient.

Fig. 5 shows a standard deviation of 0.15 for cost, 9.71 days for manufacturing ability, 0.85 USD for revenue, and 2.02% for the delivery rate. Overall, the table provides a comprehensive overview of the product's cost, manufacturing efficiency, revenue generation, and delivery performance, with a clear distinction between different versions and a sense of the overall trends and variability in the data.

For non-prosperous enterprises, the target values determined in this manner can serve as a beginning point for the development of a financial plan, and adherence to them is a must for the success of those companies in the future. Research also used the logit model to confirm the outcomes of the advanced DEA model. Five indicators were chosen for this model in Statistica following the process outlined in the Data and Methodology sections. The logit model parameters are estimated in Table V. Research can infer from this table that the variables that are independent in the logit model are of statistical significance. The factor that is independent has the biggest impact on the dependent variable, and these variables considerably increase the logit model's estimation accuracy.

TABLE V. LOGISTIC COEFFICIENT FUNCTION

Variables	Estimate	SD	p-value
NITA	-40.9545	12.48771	0.011041
WCTA	-6.1531	1.89539	0.001170
EBIE	-1.4029	0.50668	0.005628
ED	0.0654	0.13257	0.622264
CLTA	-1.0044	1.58715	0.525876

The Table V and Fig. 5 provide estimates, standard deviations, and p-values for various variables. The variable NITA has an estimated coefficient of -40.9545, with a standard deviation of 12.48771, and a p-value of 0.011041. Similarly, WCTA has an estimated coefficient of -6.1531, a standard deviation of 1.89539, and a p-value of 0.001170. The variable EBIE has an estimated coefficient of -1.4029, a standard deviation of 0.50668, and a p-value of 0.005628. On the other hand, the variable ED has an estimated coefficient of 0.0654, a standard deviation of 0.13257, and a relatively high p-value of 0.622264. Lastly, the variable CLTA has an estimated coefficient of -1.0044, a standard deviation of 1.58715, and a

p-value of 0.525876. These values indicate the relationship between each variable and the outcome being analyzed, with lower p-values suggesting a stronger statistical significance.

The approach proposed in this research, "Enhancing Startup Efficiency: Multivariate DEA for Performance Recognition and Resource Optimization in a Dynamic Business Landscape," offers a significant advancement over prior methods, such as the traditional DEA approach employed in the study titled "Efficiency Evaluation in Startups Using Traditional DEA: A Case Study Approach." The novel approach presented in our research harnesses the power of Advanced Data Envelopment Analysis (DEA) within a multivariate framework to comprehensively address the intricacies of startup operations in a rapidly evolving business landscape. Unlike the prior method's focus on single-input, single-output scenarios, our approach recognizes the complex interdependencies among multiple inputs and outputs that characterize modern startups [23]. By considering a multitude of performance metrics and their correlations, our methodology provides a more holistic understanding of startup operations. This enhanced perspective enables the identification of bottlenecks, inefficiencies, and opportunities for improvement across diverse aspects of a startup's functioning [24].

Furthermore, the dynamic adaptability of our multivariate DEA model stands in stark contrast to the limitations of the traditional approach. Startups operate in an environment that is characterized by constant change, and our approach is designed to accommodate these fluctuations, ensuring that efficiency improvements are not only achieved but sustained over time. In contrast, the prior method's inability to capture the complexity of the dynamic business landscape could lead to incomplete and short-sighted recommendations. Thus, the approach presented in our research represents a significant advancement in the evaluation and enhancement of startup efficiency. Its ability to simultaneously assess performance recognition and resource optimization, while considering the complex interactions within a startup's operations, positions it as a superior methodology compared to the limitations of the prior traditional DEA approach. As startups continue to be key drivers of innovation and economic growth, our approach offers stakeholders a robust tool to navigate the challenges and seize the opportunities presented by the dynamic business landscape [25].

VI. CONCLUSION

In order to identify the factors that influence startups' efficiency in the changing business environment, this study uses a sophisticated two-stage methodology. Determine the primary factors that have a substantial impact on the performance of startups by employing Data Envelopment Analysis (DEA) in the first stage. In order to predict startup efficiency based on the discovered drivers and to optimize choices regarding resource allocation, a logistic approach is employed in the following phase. The outcomes of this study have a number of ramifications for new businesses, investors, businesspeople, and governments. Startups could strategically utilize their resources to maximize efficiency and overall performance by understanding the relationships between the

determinants and performance. This aids in risk reduction and maximizes the use of scarce resources. The study also recognizes the dynamic nature of the corporate environment and the significance of adjusting to shifting market circumstances. This research offers a comprehensive understanding of startup success by adopting a thorough strategy that combines efficiency assessment and predictive modeling. Beyond specific startups, this study's ramifications are wide-ranging. Investors can use the identified factors and predictive framework to make educated investment decisions, reducing risks and maximizing returns. Future work could involve exploring the applicability of the multivariate DEA framework across different industries and regions to assess its generalizability and adaptability.

REFERENCES

- [1] G. A. Knight and S. T. Cavusgil, "Innovation, organizational capabilities, and the born-global firm," *Journal of international business studies*, vol. 35, pp. 124–141, 2004.
- [2] F. Caputo, D. Magni, A. Papa, and C. Corsi, "Knowledge hiding in socioeconomic settings: matching organizational and environmental antecedents," *Journal of Business Research*, vol. 135, pp. 19–27, 2021.
- [3] M. Del Giudice, M. R. Della Peruta, and V. Maggioni, "A model for the diffusion of knowledge sharing technologies inside private transport companies," *Journal of Knowledge Management*, vol. 19, no. 3, pp. 611–625, 2015.
- [4] H. Inkinen, "Review of empirical research on knowledge management practices and firm performance," *Journal of knowledge management*, 2016.
- [5] E. Battisti, S. Alfiero, R. Quaglia, and D. Yahiaoui, "Financial performance and global start-ups: the impact of knowledge management practices," *Journal of International Management*, vol. 28, no. 4, p. 100938, 2022.
- [6] D. Tandon, K. Tandon, and N. Malhotra, "An evaluation of the technical, pure technical and scale efficiencies in the Indian banking industry using data envelope analysis," *Global Business Review*, vol. 15, no. 3, pp. 545–563, 2014.
- [7] S. K. Roy, R. L. Gruner, and J. Guo, "Exploring customer experience, commitment, and engagement behaviours," *Journal of Strategic Marketing*, vol. 30, no. 1, pp. 45–68, 2022.
- [8] S. U. Kucuk, "Macro-level antecedents of consumer brand hate," *Journal of Consumer Marketing*, vol. 35, no. 5, pp. 555–564, 2018.
- [9] M. E. Adamseged and P. Grundmann, "Understanding Business Environments and Success Factors for Emerging Bioeconomy Enterprises through a Comprehensive Analytical Framework," *Sustainability*, vol. 12, no. 21, p. 9018, Oct. 2020, doi: 10.3390/su12219018.
- [10] A. Arabmaldar, B. K. Sahoo, and M. Ghiyasi, "A generalized robust data envelopment analysis model based on directional distance function," *European Journal of Operational Research*, 2023.
- [11] H. Fukuyama, M. Tsionas, and Y. Tan, "Dynamic network data envelopment analysis with a sequential structure and behavioural-causal analysis: Application to the Chinese banking industry," *European Journal of Operational Research*, vol. 307, no. 3, pp. 1360–1373, 2023.
- [12] W.-K. K. Hsu, S.-H. S. Huang, and N. T. Huynh, "An assessment of operating efficiency for container terminals in a port—An empirical study in Kaohsiung Port using Data Envelopment Analysis," *Research in Transportation Business & Management*, vol. 46, p. 100823, 2023.
- [13] A. Yu, Y. Shi, J. You, and J. Zhu, "Innovation performance evaluation for high-tech companies using a dynamic network data envelopment analysis approach," *European Journal of Operational Research*, vol. 292, no. 1, pp. 199–212, Jul. 2021, doi: 10.1016/j.ejor.2020.10.011.
- [14] E. Lafuente, L. Szerb, and Z. J. Acs, "Country level efficiency and national systems of entrepreneurship: a data envelopment analysis approach," *The Journal of Technology Transfer*, vol. 41, pp. 1260–1283, 2016.
- [15] M. Kapelko, "Evaluating efficiency in the framework of resource-based view of the firm," Evidence from Polish and Spanish textile industry. Departamento de Economia de la Empresa, Universitat Autònoma de Barcelona (mimeo)(http://selene.uab.es/dep-economia-empresa/Jornadas/Papers/2005/KapelkoMagdalena_PaperIIJornaPreCongres.pdf), 2005.
- [16] Anderson and Stejskal, "Diffusion Efficiency of Innovation among EU Member States: A Data Envelopment Analysis," *Economies*, vol. 7, no. 2, p. 34, Apr. 2019, doi: 10.3390/economies7020034.
- [17] G. Liang, L. Xu, and L. Chen, "Optimization of Enterprise Labor Resource Allocation Based on Quality Optimization Model," *Complexity*, vol. 2021, pp. 1–10, Apr. 2021, doi: 10.1155/2021/5551762.
- [18] A. B. J. Ali, F. B. Ismail, Z. M. Sharif, and N. Majeed, "The organizational culture influence as a mediator between training development and employee performance in Iraqi Academic sector: University of Middle Technical," *Journal of Contemporary Issues in Business and Government Vol*, vol. 27, no. 1, pp. 1–10, 2021.
- [19] G. Xu and Z. Zhou, "Assessing the efficiency of financial supply chain for Chinese commercial banks: a two-stage AR-DEA model," *Industrial Management & Data Systems*, vol. 121, no. 4, pp. 894–920, 2021.
- [20] G. R. Amin and M. Toloo, "Finding the most efficient DMUs in DEA: An improved integrated model," *Computers & Industrial Engineering*, vol. 52, no. 1, pp. 71–77, 2007.
- [21] R. J. Vanderbei and others, *Linear programming*. Springer, 2020.
- [22] S. Haider and P. P. Mishra, "Does innovative capability enhance the energy efficiency of Indian Iron and Steel firms? A Bayesian stochastic frontier analysis," *Energy Economics*, vol. 95, p. 105128, 2021.
- [23] S. Singh, V. K. Tayal, H. P. Singh, and V. K. Yadav, "Design of PSO-tuned FOPI & Smith predictor controller for nonlinear polymer electrolyte membrane fuel cell," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–22, 2022.
- [24] Z. Zhang, Y. Xiao, Z. Fu, K. Zhong, and H. Niu, "A study on early warnings of financial crisis of Chinese listed companies based on DEA-SVM model," *Mathematics*, vol. 10, no. 12, p. 2142, 2022.
- [25] Y. Li et al., "bSRWPSO-FKNN: A boosted PSO with fuzzy K-nearest neighbor classifier for predicting atopic dermatitis disease," *Frontiers in Neuroinformatics*, vol. 16, p. 1063048, 2023.

Design and Improvement of New Industrial Robot Mechanism Based on Innovative BP-ARIMA Combined Model

Yuanyuan Liu

Shanxi Polytechnic College, Taiyuan, Shanxi, 030006, China

Abstract—The main innovation of Industry 4.0, which involves human-robot cooperation, is transforming industrial operation facilities. Robotic systems have been developed as modern industrial solutions to assist operators in carrying out manual tasks in cyber-physical industrial environments. These robots integrate unique human talents with the capabilities of intelligent machinery. Due to the increasing demand for modern robotics, numerous ongoing industrial robotics studies exist. Robots offer advantages over humans in various aspects, as they can operate continuously. Enhanced efficiency is achieved through reduced processing time and increased industrial adaptability. When deploying interactive robotics, emphasis should be placed on optimal design and improvisation requirements. Robotic design is a very challenging procedure that involves extensive development and modeling efforts. Significant progress has been made in robotic design in recent years, providing multiple approaches to address this issue. Considering this, we propose utilizing the Backpropagation Autoregressive Integrated Moving Average (BP-ARIMA) combination model for designing and improving a novel industrial robot mechanism. The design outcomes were evaluated based on performance indicators, including accuracy, optimal performance, error rate, implementation cost, and energy consumption. The evaluation findings demonstrate that the suggested BP-ARIMA model offers optimal design for industrial robotics.

Keywords—Industry 4.0; robotics; design; backpropagation autoregressive integrated moving average (BP-ARIMA); and operation facilities

I. INTRODUCTION

An industrial robot is a revolutionary machine designed to ease the burden of repetitive factory tasks. Assembly plants are examples of highly dynamic settings that have significantly benefited from this invention. These robots are installed as fixed, imposing features of the factory space, with various other workers' activities revolving around them [1]. Industrial robots are movable platforms with sensors, processors, and actuators that can function autonomously. These systems equip robots to perform discrete operations inside elaborate processing or production pipelines. These devices have three or more axes of motion and may be programmed to carry out various tasks, which is why they are sometimes referred to as robotic systems [2]. Multiple mechanisms power these automated systems; the most common are electric motors, hydraulic drives, and pneumatic controls. In terms of price-to-performance, electric motors are the way to go because of their

reliable power source and straightforward construction. Their increasing popularity may be attributed to the wide variety of jobs they can do, which includes welding joints, selecting and arranging things, piercing, and sawing. In addition, their effectiveness exceeds that of traditional propulsion methods. Fig. 1 shows some of the many ways industrial robotics technology is being used, demonstrating the far-reaching impact of robotics [3].

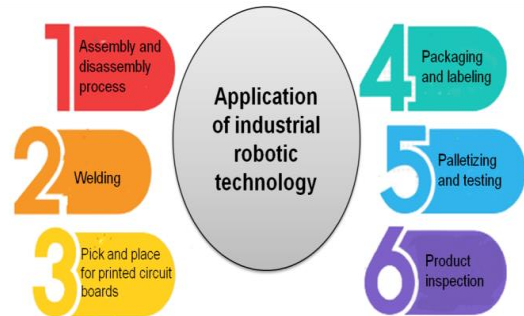


Fig. 1. Application of industrial robotics technology.

These robots also greatly aid monitoring efforts for controlling industrial processes and assuring product quality. The efficiency of future factories and enterprises will significantly benefit from their influence. Robotic calibration aims to discover statistical approaches that more accurately represent the machine's capabilities, taking precision beyond the level of nominal kinematic models. The price tag is manageable for all the quality gains we're making. Manipulators, end effectors, input gear, controllers, and locomotion systems are the fundamental building blocks of every effective industrial robot [4]. As a measure of technical progress, the degree of automation achieved may be defined by the extent of its integration. While robots have come a long way in intelligence, they still often need human aid to deal with unpredictable workloads and environmental conditions. Precision and endurance are where robots thrive, but humans have the upper hand when it comes to things like intuition, responsiveness, and flexibility. This collaboration makes the most of each party's capabilities. As the field of autonomous systems continues to grow and adapt to new applications and problems, route planning has become more important [5]. There are unique challenges and necessities for industrial robot systems as they become more flexible. Traditional online training and offshore procedures are two prominent uses for industrial robots. These methods are constantly developing to

keep up with the increasing complexity of today's projects and software. Industrial robots are on the cusp of altering the assembly line and the global economy. These robots are set to spearhead this shift because of their rising intelligence, mobility, interaction, and adaptability capabilities. Even while people are still in the driver's seat in intelligent factories, automation is closing the gap [6].

Energy-efficient methods, such as green manufacturing, are in high demand because of rising worldwide concerns about pollution and resource depletion. Techniques that reduce pollution and waste during production are emphasized [7]. Automation and robotization, in which machines, rather than humans, do laborious, repetitive activities, are becoming more common. Industrial robots can currently perform various tasks, mimicking human agility and productivity. Unlike humans, they never tire, giving them a distinct edge [8]. Researchers have shown that robotization has significant advantages for enterprises, including higher productivity, lower manufacturing costs, and better utilization. Industrial robots increasingly perform machining processes, including chamfering, deburring, grinding, and polishing.

Compared to conventional machining centers, they provide many benefits, including adaptability, enough space for work, and low initial investment costs. Industrial robot efficiency is measured by creating different control systems and paths. High-speed cameras monitor mobility information, assessing velocity, acceleration, and direction accuracy [9]. A laser tracker-based adjustment method is offered in the pursuit of precision. This method detects residual errors through tool location measurements, enhancing motion planning and ensuring obstacle-free operations—the trajectory of robotization points toward the proliferation of industrial robots in various sectors. According to projections by the Boston Consulting Group, a global automation revolution is on the horizon as industries approach the point where robotization becomes economically feasible [10].

As for the remainder of the paper, Section II provides the relevant research and suggested approaches. Section III provides a description, Section IV displays the findings, and Section V shows the final results of the article.

II. RELATED WORKS

The study [11] suggests a technique based on innovation performance standards that can conduct a topology optimization for industrial robots applicable across the board in the workplace or while using a specific set of pathways in Industry 4.0. The best robot designs or trajectories for which extreme performance will be reached are calculated and repeatedly modified to require the selected efficiency indicators to be applicable worldwide. It uses the elastic models' structural features to lower the computational burden of these performance metrics and thereby lessen the computing time needed to calculate them. The Linearization Method, our last optimization technique, produces results in computing time comparable to conventional topological optimization algorithms. Still, its implementation is more straightforward, making it simple to do customization or enhancement. The study [12] analyses industrial robot control performance and dependability while considering the impact of unknown

factors. Initially, the Denavit-Hartenberg technique is used to create the kinematic models of industrial robots, which assumes the connection extents and component rotational degrees to be unknowable factors. To do the durability study, the sensitive factors are identified using the Sobol approach, which is also used to examine the sensitivity of unknown factors for the strategizing accuracy of industrial robots. The research [13] thoroughly analyzes chatter-related concerns that arise throughout robotic machining activities, covering processes, mitigating tactics, and techniques for identifying regeneration chatter and mode coupling chatter. Industry robots' weak stiffening and relationship design may cause regeneration and talk-in-phase couplings under various cutting circumstances. A set of recommendations are offered to assist in differentiating between the double chatter mechanisms utilized in robotic machining after a comparison of the two is conducted.

The research [14] proposes an intelligent failure detection method for multi-joint industrial robots based on attitude data. The attitude change of the final joint is used to represent the transmission defect of robot components based on the study of the transmission mechanism. The multi-joint robot's last joint is equipped with only one attitude sensor as part of an affordable data-collecting method. The study [15] outlines using the virtual reality (VR) digital twin of physical architecture to analyze how people respond, including predictable and random robot movements. Human responses are analyzed, and the VR environment's efficiency is validated using various existing measures and a newly created Kinetic Energy Ratio meter. It has made it more difficult for governments to enact laws governing whether people and machines must coexist near one another, as well as the development of human-robot collaboration tactics. The research [16] offers the first thorough analysis of how the employment of industrial robots affects the emissions intensity of production. It discovered that industrial robots might significantly increase the energy intensity of production, and our hypothesis survived several robustness tests. Also, the technical advancement impact and complementing effect among workers and industrial robots play a role in these economic benefits. Lastly, they discovered a complex relationship between production emission intensity and industrial robots. Instead of sustainable energy intensity, robotic systems may impact non-renewable energy intensity. The author in [17] presents a method for analyzing how people respond to both expected and unanticipated robot movements. It employs an augmented actuality digital twin of a physical structure.

Several standard measures and a newly created Kinetic Energy Ratio metric are employed to analyze user responses and confirm the efficacy of the VR environment. It has made it more difficult for governments to enact laws governing how people and robots should coexist near one another, as well as the development of human-robot collaboration tactics. In the study [18], 460 senior managers and owners of ACMCs in India were surveyed on their intentions to adopt and plans to deploy InRos in their organizations. 4.0 Industry Compatibility is one of the critical factors influencing InRos adoption intention, according to the study, external pressure, perceived advantages, and vendor support. Interestingly, the study also

shows that official backing and IT infrastructure have little impact on a person's decision to embrace InRos. The data further reveals that perceived cost concerns adversely affect the association between adoption intention and probable InRos usage in ACMCs. The research contributes to the theory by using the conventional TOE framework and finding, counterintuitively, that Indeed facilities are not a primary factor of technology acceptance.

III. MATERIALS AND METHOD

Robot implementations have become prominent because of the industrial sector's increasing demand for versatility, cost-effectiveness, and performance. Industrial robots combine a workspace and collaborate with human employees in these settings. For industrial robotics design, the BP-ARIMA is recommended.

Backpropagation Autoregressive Integrated Moving Average (BP-ARIMA) for industrial robotic design

Backpropagation combines the precision of ARIMA models with the adaptability of neural networks, Autoregressive Integrated Moving Average (BP-ARIMA) performs very well when applied to robotics. This integration may allow more accurate modeling and prediction of robotic systems' dynamic behaviors and time-varying patterns. Foreseeing these nonlinear correlations and complex linkages is essential to robotic operations. BP-ARIMA's use of neural networks for feature extraction and the backpropagation approach for iterative refinement makes this possible. BP-ARIMA's adaptability makes it useful for various robotic applications, including defect detection, efficiency enhancement, and sensor data prediction.

The high nonlinear adaptation capability of backpropagation (BP) has resulted in its widespread application in several prediction disciplines. In dependability design, this approach is relatively uncommon. As a result, a reliable robotic design technique based on BP has been established. The intake layer, concealed layer, and output layer are the three layers that compose a BP neural network. The incoming layers are implicitly linked to the output nodes by concealed neurons. The link between two neurons' weight attributes reflects the relationship's degree. Establishing a BP model entails choosing the number of neurons for the hidden layer, the number of source and outcome neurons, and the weight ratios of the linkages. The BP algorithm's principle may be concisely explained as follows. The output is produced in the first phase when the inputs spread outward. The discrepancy between the created and actual results is determined in the second stage, transmitted back to the entry layer, and the connection lifts are modified to lower the error. Once the resultant network resembles the trained data sufficiently, that is, till the errors between the expected and actual outcomes are suitably modest, this procedure of modifying the connectivity weights is maintained.

A BP system is implemented by identifying the number of neurons for the input and output layers, finding out how many neurons are in the concealed layer, and figuring out the weights of the connections. There are three input criteria, and their related inputs are the industrial model, optimized energy, and

efficient performance. There is just one outgoing neuron that, under various circumstances, relates to the optimal performance design. The following equation 1 may be used in designing even if there is no chance to precisely compute the ideal number of concealed layer neurons depending on data.

$$c = \sqrt{i + o + a} \quad (1)$$

Where a is steady between 1 and 10, c is the quantity of concealed layer neurons, i is the number of incoming layer neurons, o is the number of output layer neurons, and a is changed to reduce the prediction error to attain the optimum matching effectiveness. During the training phase, the weight levels of linkages are continuously modified until the design inaccuracies are decreased under a predefined level. The symmetric randomized design approach, which has homogenous distribution and fair evaluation, is used to choose the training data.

$$b_i = \frac{a_i - \min a_i}{\max(a_i) - \min(a_i)} \quad (2)$$

The preceding part provided the training design information. The sampling information is standardized using equation 2 to assure resolution; $\max(a_i)$ and $\min(a_i)$ denote the greatest and lowest values within the group of the i -th input. a_i and b_i denote the raw and standardized data of the i -th input, correspondingly. The information is re-ranged between 0 and 1 after the standardization procedure. The BP incorporates two processes, data forward propagating and erroneous backward propagating, both dependent on gradient reduction. While the system is training, the data is sent from the intake layer to the output layer. If the outcomes do not match the desired results, the gradient is sent back into the system to modify the load and skew of each neuron and reduce the error between anticipated and actual data. Equation 3 describes the BP's goal functionality.

$$G = \frac{1}{2 \times a} \sum_{i=1}^N \sum_{j=1}^L x \left(\frac{x_{ij} - \hat{x}_{ij}}{a} \right)^2 \quad (3)$$

where a is the learning sample amount, L is the output variable's size, and x_{ij} and \hat{x}_{ij} are the actual result and estimation data, correspondingly. Equation 4 defines the system x_k 's output value.

$$x_k = f(z_{1k} \times o_1 + \dots + z_{jk} \times o_j + \dots + z_{nk} \times o_n + a_k)$$

o_j is the output level of the j th concealed layer neurons, z_{jk} is the weighted connecting the j th hidden layer nodes and the k th output layer neurons, and a_k is the biased value of the k th output layer nodes and stands for the activating factor. The networking design weights and biases must be changed following the training inaccuracy for the output values to be near the intended value. Equation 4 and 5 modifies the value upgrade equation for concealed layers and distortions.

$$z'_{jk} = w_{jk} - \frac{a}{n} \sum_{k=1}^n \Delta v - o_j \quad (4)$$

$$a'_j = a_j - \frac{1}{n} \sum_{k=1}^n \Delta v \quad (5)$$

Where n is the total amount of nodes in the output layer, w'_{jk} is the updated weights connecting the nodes in the j th concealed layer and the k th output units, a'_j and n is the modified

bias of the k th output layer nodes. V is the learning variance, which was adjusted to be $\Delta v = x_k - \hat{x}_k$.

Unfortunately, the numerical accuracy frequently falls short of the standards. Numerous academics enhance the BP variables to increase the design reliability of BP.

Autoregressive Integrated Moving Average (ARIMA) is among the greatest widely used procedures for designing optimal robots. It is a flexible approach that can accommodate different time series behaviors. The essential need is for the information to be stable, meaning they maintain their statistical features throughout the process. Differentiation and other nonlinear transitions, such as the logistic function, can convert non-stationary series into static ones. The ARIMA functions as a filter that tries to isolate the signal from the earlier noise using the signal framework to enhance it. The term ARIMA comprises three parts that are united to make the model. The Autoregressive (AR) portion is the first component. This section aims to demonstrate the effect of earlier data. It is executed by regression using the most recent p values from the series. Equation 6 serves as the model for AR.

$$\hat{c}_k = s + \sum_{j=1}^p \phi_j c_{k-j} + \epsilon_k \quad (6)$$

When s is steady, ϕ_j is a modeling variable used to weigh prior variables and an arbitrary inaccuracy. Incorporation is the second element (I). By comparing the information by level d , it achieves its goal of making it stable. The final component is the Moving Average (MA), which eliminates arbitrary information fluctuations and retrieves value from the prior inaccuracy components. The MA of the most recent p predicted inaccuracies is used to produce it. Equation 7 provides the MA system.

$$\hat{c}_k = \mu_k + \sum_{i=1}^p \theta_j \epsilon_{k-j} + \epsilon_k \quad (7)$$

Where is the series' average up to moment k , $k = \hat{c}_k - c_k$ is the prediction's inaccuracy in the prior, and j is a modeling variable that accounts for previous mistakes. The entire ARIMA structure is shown in Equation 8, where c is the differenced series created by combining Equations 7 and 8. The three variables, p , d , and q , may be changed to emphasize some components more than others. With various settings, many prototypes that could be used with multiple series may be produced.

$$\hat{c}_k = s + \sum_{j=1}^p \phi_j c'_{k-j} + \sum_{i=1}^p \theta_j \epsilon_{k-j} + \epsilon_k \quad (8)$$

Likewise, efficient robotic design is incorporated into the industrial sector.

IV. RESULT AND DISCUSSION

1) *Performance analysis*: This study introduces a brand-new design measure built on BP-ARIMA to enhance the industrial sector and improve robot design. In this section, the evaluation is discussed. Accuracy, optimal performance, error rate, implementation cost, and energy consumption are used to evaluate the effectiveness of the suggested system. The existing techniques used for comparison are the hybrid Grasshopper optimization algorithm and Nelder–Mead

(HGOANM) [19], fuzzy wavelet neural networks (RFWNNs) [20], and radial basis function neural network (RBFNN) [21].

A. Accuracy

In industrial robots, accuracy and repeatability are essential. The robot's ability to reliably go to a given spot is measured by its repeatability. The difference between the position the robot is planned to reach and the value of the work the robot arrives at is known as accuracy. Analyzing the design accuracy of time intervals provides information about the performance of the suggested framework. Fig. 2 indicates the accuracy of the proposed method. The accuracy outcome of the recommended way is shown in Table I.

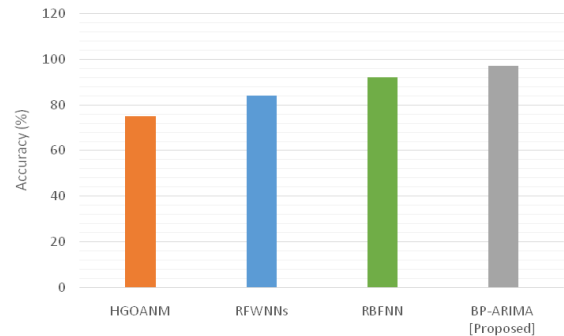


Fig. 2. Accuracy of the proposed and existing techniques.

TABLE I. OUTCOME OF ACCURACY

Methods	Accuracy (%)
HGOANM	75
RFWNNs	84
RBFNN	92
BP-ARIMA [Proposed]	97

B. Optimal Performance

Optimal performance describes an optimal state where individuals are wholly absorbed in the activity. Industrial robots use optimization to locate the most effective method for enhancing 3D space accuracy, decreasing vibrations, selecting ideal robot base points for applications to cut down on required times, and discovering creative or operational factors that guarantee lower energy consumption. Fig. 3 suggests the optimal performance of the proposed method. The outcome of the optimal performance recommended method is shown in Table II. It shows the suggested approach is more Optimal than the existing approach.

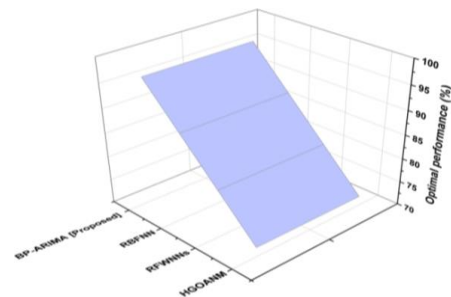


Fig. 3. Optimal performance of the proposed and existing techniques.

TABLE II. THE OUTCOME OF OPTIMAL PERFORMANCE

Methods	Optimal performance (%)
HGOANM	72
RFWNNs	80
RBFNN	88
BP-ARIMA [Proposed]	96

C. Error Rate

The error rate measures how much a model deviates from the genuine model in its predictions. For design techniques, the phrase error rate is often used. The error rate gauges how far a model's improvement strays from reality. The error rate of a sector is the percentage of operational errors made by that sector. Given the expense of correcting errors, manufacturing errors should be avoided at all costs. By dividing the overall amount of incorrect predictions on the testing sample by every one of the statements on the testing dataset, on the other hand, it is possible to get the error rate. The recommended method's error rate is shown in Fig. 4. The results of the suggested technique are shown in Table III. It demonstrates how the recommended approach has a lower mistake rate than the existing approach.

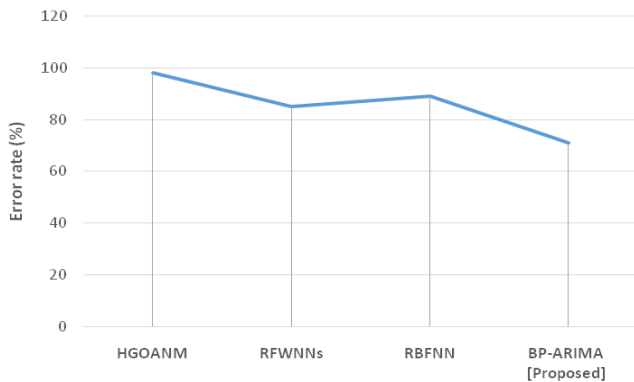


Fig. 4. Error rate of the proposed and existing techniques.

TABLE III. THE OUTCOME OF THE ERROR RATE

Methods	Error rate (%)
HGOANM	98
RFWNNs	85
RBFNN	89
BP-ARIMA [Proposed]	71

D. Implementation Cost

Manufacturers may use the cost of quality to evaluate and enhance their quality requirements. In contrast to the expenses related to inner and outer breakdowns, it is a technique for identifying and quantifying when most of an organization's resources are spent on prevention and maintaining product quality. Implementation costs are those associated with planning and carrying out a strategy for implementing particular or much particular proof treatment. These factors include labor, power, resources, life-long process maintenance, and manufacturing inputs to operate a robot properly.

According to the company sector and scale of the operation, these expenses differ wildly because of the various kinds of production facilities. The recommended method's implementation cost is shown in Fig. 5. The implementation cost results of the suggested technique are shown in Table IV. It proves that the proposed method uses less cost.

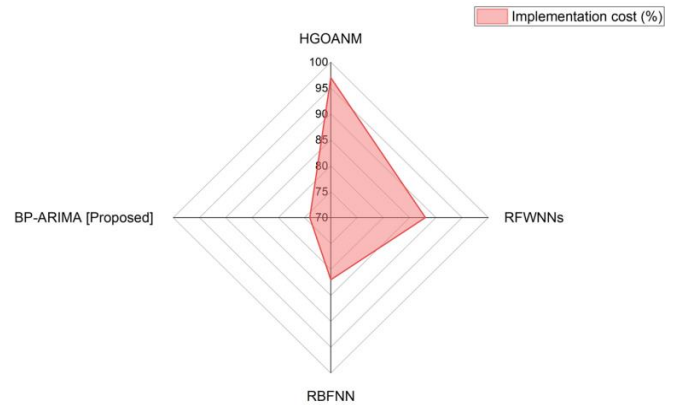


Fig. 5. Implementation cost of the proposed and existing techniques.

TABLE IV. THE OUTCOME OF THE IMPLEMENTATION COST

Implementation cost (%)	Implementation cost (%)
HGOANM	97
RFWNNs	88
RBFNN	82
BP-ARIMA [Proposed]	74

E. Energy Consumption

The controllers, conditioning air, engine, and friction at the robotic connection are some parts of the robotic system that use energy. Energy and other sources like gasoline engines or compressed gasses may be used to power action. Electric actuators are most often used in smaller, interior robotics of the kind that the beginning constructor is far more able to design. A comprehensive and realistic industrial robot model using less energy consumption has been presented. The recommended method's energy consumption is shown in Fig. 6. The energy consumption outcome of the suggested technique is shown in Table V.

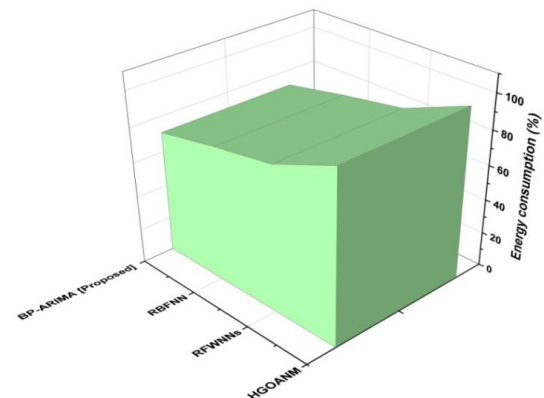


Fig. 6. Energy consumption of the proposed and existing techniques.

TABLE V. OUTCOME OF THE ENERGY CONSUMPTION

Methods	Energy consumption (%)
HGOANM	99
RFWNNs	84
RBFNN	78
BP-ARIMA [Proposed]	71

V. DISCUSSION

The HGOANM may take more work to implement and comprehend due to the multiple optimization methodologies it incorporates. Due to its sensitivity to beginning circumstances, the Nelder-Mead method's convergence rate may be sluggish for high-dimensional or non-convex problems. Because of the complexity of their construction, some of the interpretability gained by using fuzzy logic in conjunction with neural networks may be lost when using fuzzy wavelet neural networks. Additional challenges that restrict the efficacy of radial basis function neural networks in high-dimensional environments include their inability to scale and the possibility of underfitting or overfitting. The BP-ARIMA model combines the ARIMA models' predictive ability and neural networks' flexibility. This combination accurately models and predicts complex time series data while considering nonlinear linkages and temporal interdependence. Financial forecasting, environmental monitoring, and industrial processes may benefit from BP-ARIMA's capacity to automatically identify significant characteristics and employ backpropagation for iterative learning to manage dynamic behaviors and changing patterns. BP-ARIMA implementation may need more processing resources than regular ARIMA models due to the neural network component, and its performance depends on parameter variation and training data quality. Consider these parameters to increase the model's predictive ability. Technology may reduce jobs, rising inequality, and unemployment. Robots and humans value work safety. Risk assessments and safety measures avoid harm. Teamwork robots need instruction. Mistakes and inefficiencies lower production and quality without training. Industrial human-robot cooperation requires morality, safety rules, and well-structured training.

VI. CONCLUSION

Robots are utilized more often in the workplace today to substitute people, especially for monotonous activities. The heterogeneous integration of a wide range of innovations marks industrial robot growth. It must be noted that the primary markets for industrial robots nowadays are the automobile sectors, particularly their supply networks. This indicates that a significant portion of the advancement of robots is driven by the needs arising from this production process. Thus, most robots nowadays are ideally suited to adaptable, high-volume, cost-conscious manufacturing in a highly dynamic context. This has forced robot makers to put a lot of work into meeting the fundamental standards for cost-effectiveness, high dependability, and efficiency. The creation of an industry-specific optimum design was the goal of this research. This study presents the Backpropagation Auto-Regressive Integrated Moving Average (BP-ARIMA) as an efficient method. The traditional system is evaluated and

compared for accuracy, optimal performance, error rate, implementation cost, and energy consumption. The results show that the suggested method offers an improvement and effective design for industrial robots. The performance of the proposed system may be increased in the future by using optimization techniques. Combining the BP-ARIMA model with creating a novel mechanism has several prospects for industrial robotics. When applied to industrial robots, BP-ARIMA provides a game-changing method for enhancing performance, adaptability, and prediction.

FUNDING STATEMENT

Supported by Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2020L0768)

REFERENCES

- [1] E.Fosch-Villaronga, and C. Millard, "Cloud robotics law and regulation: Challenges in the governance of complex and dynamic cyber-physical ecosystems," *Robotics and autonomous systems*, 119, pp.77-91,2019.
- [2] C.K. Lo, C.H. Chen, and R.Y.Zhong, "A review of the digital twin in product design and development," *Advanced Engineering Informatics*, 48, p.101297, 2021.
- [3] G. Zhao, P. Zhang, G.Ma, and W.Xiao, "System identification of the nonlinear residual errors of an industrial robot using massive measurements," *Robotics and Computer-Integrated Manufacturing*, 59, pp.104-114, 2019.
- [4] Z.M.Bi, C.Luo, Z.Miao, B.Zhang, W.J. Zhang, and L. Wang, "Safety assurance mechanisms of collaborative robotic systems in manufacturing," *Robotics and Computer-Integrated Manufacturing*, 67, p.102022,2021.
- [5] H. Zhang, Y.Wang, J.Zheng, and J.Yu, "Path planning of industrial robot based on improved RRT algorithm in complex environments," *IEEE Access*, 6, pp.53296-53306, 2018.
- [6] A.D. Pham, and H.J.Ahn, "High precision reducers for industrial robots driving 4th industrial revolution: state of arts, analysis, design, performance evaluation and perspective," *International Journal of precision engineering and manufacturing-green technology*, 5, pp.519-533, 2018.
- [7] P.Barosz, G. Gołda, and A.Kampa, "Efficiency analysis of manufacturing line with industrial robots and human operators," *Applied Sciences*, 10(8), p.2862, 2020.
- [8] J.H. Jung, and D.G. Lim, "Industrial robots, employment growth, and labor cost: A simultaneous equation analysis," *Technological Forecasting and Social Change*, 159, p.120202, 2020.
- [9] F.Gao, W.Tang, J.Huang, and H.Chen, "Positioning of Quadruped Robot Based on Tightly Coupled LiDAR Vision Inertial Odometry," *Remote Sensing*, 14(12), p.2945, 2022.
- [10] K. Wu,C.Krewet, and B.Kuhlenkötter, "Dynamic performance of industrial robot in corner path with CNC controller," *Robotics and Computer-Integrated Manufacturing*, 54, pp.156-161, 2018.
- [11] S.Briot, and A.Goldsztejn, "Topology optimization of industrial robots: Application to a five-bar mechanism," *Mechanism and Machine Theory*, 120, pp.30-56, 2018.
- [12] A. Gasparetto, and L.Scalera, "From the Unimate to the Delta robot: the early decades of Industrial Robotics. In *Explorations in the History and Heritage of Machines and Mechanisms: Proceedings of the HMM IFToMM Symposium on History of Machines and Mechanisms* (pp. 284-295)," Springer International Publishing, 2019.
- [13] L.Yuan, Z.Pan, D.Ding, S.Sun, and W.Li, "A review on chatter in robotic machining process regarding both regenerative and mode coupling mechanism," *IEEE/ASME Transactions on mechatronics*, 23(5), pp.2240-2251, 2018.
- [14] J.Long, J.Mou, L. Zhang, S.Zhang, and C.Li, "Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint

- industrial robots,” *Journal of manufacturing systems*, 61, pp.736-745, 2021.
- [15] H. Lu, M. Du, k. Qian, X. He and k. Wang. GAN-based data augmentation strategy for sensor anomaly detection in industrial robots. *IEEE Sensors Journal*, 22(18), pp.17464-17474, 2021.
- [16] E.Z.Wang, C.C. Lee, and Y.Li, “Assessing the impact of industrial robots on manufacturing energy intensity in 38 countries,”*Energy Economics*, 105, p.105748,, 2022.
- [17] J.O.Oyekan, W.Hutabarat, A.Tiwari, R.Grech, M.H. Aung, M.P.Mariani, L.López-Dávalos, T.Ricaud, S.Singh and C.Dupuis, “The effectiveness of virtual environments in developing collaborative strategies between industrial robots and humans,” *Robotics and Computer-Integrated Manufacturing*, 55, pp.41-54, 2019.
- [18] R.Pillai, B.Sivathanu, M.Mariani, N.P. Rana, B.Yang, and YK Dwivedi “Adoption of AI-empowered industrial robots in auto component manufacturing companies,” *Production Planning & Control*, 33(16), pp.1517-1533, 2022.
- [19] B.S.Yildiz, N.Pholdee, S.Bureerat, A.R. Yildiz, and S.M. Sait, “Robust design of a robot gripper mechanism using new hybrid grasshopper optimization algorithm,”*Expert Systems*, 38(3), p.e12666,2021.
- [20] V.T. Yen, W.Y. Nan, and P.VanCuong,”Recurrent fuzzy wavelet neural networks based on robust adaptive sliding mode control for industrial robot manipulators,” *Neural Computing and Applications*, 31, pp.6945-6958, 2019.
- [21] F.Luan, J.Na, Y.Huang, and G.Gao, “Adaptive neural network control for robotic manipulators with guaranteed finite-time convergence,”*Neurocomputing*, 337, pp.153-164, 2019.

A Proposed Approach for Monkeypox Classification

Luong Hoang Huong, Nguyen Hoang Khang, Le Nhat Quynh, Le Huu Thang,
Dang Minh Canh, Ha Phuoc Sang
Department of Information Technology, FPT University, Can Tho, Viet Nam

Abstract—Public health concerns have been heightened by the emergence and spread of monkeypox, a viral disease that affects both humans and animals. The significance of early detection and diagnosis of monkeypox cannot be overstated, as it plays a crucial role in minimizing the negative impact on affected individuals and safeguarding public health. Monkeypox poses a considerable threat to human well-being, causing physical discomfort and mental distress, while also posing challenges to work productivity. This study proposes an applied model that combines deep learning models such as: ResNet-50, VGG16, MobileNet and machine learning models such as: Random Forest Classifier, K-Nearest Neighbors Classifier, Gaussian Naive Bayes Classifier, Decision Tree Classifier, Logistic Regression Classifier, AdaBoost Classifier to classify and detect monkeypox. The datasets are used in this research are the Monkeypox Skin Lesion Dataset (MSLD) and the Monkeypox Image Dataset (MID) that have total 659. Subjects range from healthy cases to severe skin lesions. The test results show that the model which combines deep learning and machine learning models achieves positive results, with Accuracy being 0.97 and F1-score being 0.98.

Keywords—Monkeypox; machine learning; deep learning; skin lesions

I. INTRODUCTION

Monkeypox is a viral zoonosis that can cause illness in humans and animals. The first recorded case of monkeypox in humans was in 1970 in the Democratic Republic of the Congo [1]. Since then, there have been several outbreaks of monkeypox, but the current outbreak is the largest and most widespread ever recorded [2]. The current outbreak of monkeypox has raised public health concerns due to its rapid spread and the potential for severe illness. Early detection and diagnosis of monkeypox is essential for minimizing the negative impact on affected individuals and safeguarding public health [3].

According to the available data [3], there have been 2891 confirmed monkeypox cases in the United States as of July 22, 2022. Globally, there have been a total of 71,237 laboratory-confirmed cases and 26 related deaths reported up to October 6th, 2022 [2]. Among the six World Health Organization regions, the Americas demonstrated the highest total laboratory-confirmed monkeypox cases (45,342 cases), followed by the European Region (24,889 cases), the African Region (727 cases), the Western Pacific Region (189 cases), the Eastern Mediterranean Region (67 cases), and the South-East Asia Region (23 cases) [2]. The nation with the highest cumulative monkeypox cases was the United States of America (26,723 cases), followed by Brazil (8,147 cases), Spain (7,209

cases), France (4,043 cases), the United Kingdom (3,654 cases), and Germany (3,640 cases) [2].

Traditional methods for detecting and diagnosing monkeypox, such as PCR testing, can be time-consuming and expensive. However, recent advancements in deep learning and machine learning models have provided an opportunity to develop an applied model for the classification and detection of monkeypox. By leveraging these technologies, the author aims to improve the efficiency and accuracy of monkeypox detection.

By combining the insights gained from the analysis of the dataset, the author can enhance the author's understanding of the current monkeypox outbreak and develop more effective strategies for its control and prevention. The proposed applied model, which integrates deep learning and machine learning models, holds promise for the timely and accurate classification and detection of monkeypox, thereby aiding in the mitigation of its impact on public health.

The test results show that the model which combines deep learning and machine learning models achieves positive results, with accuracy being 0.97 and F1-score being 0.98.

This research article has used the combination of deep learning models such as: ResNet-50, VGG16, MobileNet and machine learning models such as: Random Forest Classifier, K-Nearest Neighbors Classifier, Gaussian Naive Bayes Classifier, Decision Tree Classifier, Logistic Regression Classifier, AdaBoost Classifier to classify and detect monkeypox, with the foremost contributions of this paper are as follows:

- The author introduces an innovative and advanced solution to effectively address the challenge of detecting monkeypox.
- By combining deep learning and machine learning models, a formidable approach is devised to deliver exceptionally precise classification outcomes for monkeypox.
- Finding the best set of hyperparameters with fine-tuning.
- Remarkable headway is made in accurately classifying instances of the monkeypox disease, marking a significant stride forward in this area of research.

The article consists of five parts. Firstly, Section I serves as an introduction, providing a definition of the problem. Next, Section II presents related works that have been conducted. Moving on to Section III, the implementation method is

thoroughly explained. Section IV showcases the experimental results obtained from the research. Finally, in Section V, concluding remarks are provided to wrap up the article.

II. RELATED WORK

In this research [4], the author used these methods for data collection: Web-scraping for Image Collection, Expert Screening, Data Preprocessing, Augmentation. And using seven deep learning models named: ResNet-50, DenseNet121, Inception-V3, SqueezeNet, MnasNet-A1, MobileNet-V2, and ShuffleNet-V2-1x to conduct training on a dataset of diseases: Monkeypox, Chickenpox, Smallpox, Cowpox, Measles, Healthy; produced the best result in terms of accuracy (83%).

In addition, in [5], the author aims to compare different pre-trained deep learning (DL) models for Monkeypox virus detection on the disease dataset: Monkeypox, Chickenpox, Measles, Normal. Those deep learning models are VGG16, ResNet, Inception-V3, InceptionResNet, Xception, MobileNet, DenseNet, EfficientNet; produced the best result in average Precision, Recall, F1-score, and Accuracy of 85.44%, 85.47%, 85.40%, and 87.13% respectively.

Furthermore, in [6] researcher examines various deep convolutional neural network (CNN) models and several machine learning classifiers to detect monkeypox disease by analyzing skin images. Specifically, the study utilizes bottleneck features from three CNN models (AlexNet, GoogleNet, and Vgg16Net) in conjunction with multiple machines learning classifiers, including SVM, KNN, Naïve Bayes, Decision Tree, and Random Forest. The findings indicate that when using Vgg16Net features, the Naïve Bayes classifier achieves the highest accuracy rate of 91.11%.

Moreover, within [7], the author employs deep learning techniques to identify Monkeypox in digital images of skin. Various models including Support Vector Machines, ResNet-50, VGG16, SqueezeNet, and InceptionV3 were utilized. The skin data was acquired from Google using web-scraping techniques with Python's BeautifulSoup, SERP API, and requests libraries. The most successful model among them was VGG16, with an accuracy of 0.96 and an F1-score of 0.92.

In [8], the researcher employed several techniques to gather data, including web scraping, expert screening, data preprocessing, and data augmentation. Two deep learning models, namely AlexNet and VGG16, were utilized to train on a dataset comprising various diseases such as Monkeypox, Chickenpox, Measles, and Healthy. Notably, the VGG16 model yielded the highest accuracy, reaching 80%.

Additionally, in [9], the author the Kaggle Monkeypox Skin Lesion Dataset (MSLD) and the Monkeypox Skin Image Dataset (MSID) for their research purposes. Four deep neural networks were utilized for transfer learning, specifically Inception V3, ResNet 50 V2, MobileNet V2, and EfficientNet-B4. The MobileNet achieved superior performance compared to the other networks on the MSID dataset, achieving a balanced accuracy of 96.55%. Conversely, for the MSLD dataset, Inception V3 exhibited the most favorable metrics, achieving a balanced accuracy of 94%.

Within [10], the author aimed to integrate a well-trained deep learning (DL) algorithm and compare its performance against various other deep learning models. The disease dataset utilized in this research comprised Monkeypox and Chickenpox. The proposed convolutional neural network (CNN) model surpassed all other DL models, achieving a remarkable test accuracy of 99%. Moreover, a weighted average precision, recall, and F1 score of 99% were documented. Impressively, AlexNet demonstrated superior performance compared to other pre-trained models, achieving an accuracy of 98%. On the other hand, VGGNet, specifically VGG16 and VGG19, exhibited the lowest performance, with an accuracy of 80.00%. The ResNet-50 model attained an accuracy of 82%, while the InceptionV3 model achieved an accuracy of 89%.

Besides, in [11], the author assessed the effectiveness of five commonly used pre-trained deep learning models, namely VGG19, VGG16, ResNet-50, MobileNetV2, and EfficientNetB3. The experimental findings indicate that the MobileNetV2 model outperformed the others in terms of classification performance. It achieved an accuracy rate of 98.16%, a recall score of 0.96, a precision score of 0.99, and an F1-score of 0.98. Furthermore, when validating the model using different datasets, the MobileNetV2 model exhibited the highest accuracy of 94%.

And recently, within [12], the author introduced and assessed a revised version of the DenseNet-201 deep learning-based CNN model called MonkeyNet. By utilizing both the original and expanded datasets, this study put forward a deep convolutional neural network that demonstrated accurate identification of monkeypox disease. The accuracy achieved was 93.19% with the original dataset and 98.91% with the augmented dataset. The author employed the "MSID" dataset, which stands for "Monkeypox Skin Images Dataset."

Finally, in [13], the author utilized several custom models including MobileNetV3-s, EfficientNetV2, ResNet-50, VGG19, DenseNet121, and Xception models. Among these models, the hybrid MobileNetV3-s model, which was optimized, performed the most outstandingly. It achieved an average F1-score of 0.98, an AUC of 0.99, an accuracy of 0.96, and a recall of 0.97.

III. PROPOSED METHOD

A. Background

1) *ResNet-50*: ResNet stands for Residual Network and represents a distinctive variant of a convolutional neural network (CNN) that was first presented in the research paper titled "Deep Residual Learning for Image Recognition" in 2015. The authors of the paper, He Kaiming, Zhang Xiangyu, Ren Shaoqing, and Sun Jian, introduced this concept. CNNs are widely utilized in various computer vision applications. ResNet-50, on the other hand, is a specific instance of a convolutional neural network that consists of 50 layers, including 48 convolutional layers, one MaxPool layer, and one average pool layer. Residual neural networks are a type of artificial neural network (ANN) that constructs networks by

assembling residual blocks [14]. The deep residual learning framework of ResNet is shown in Fig. 1.

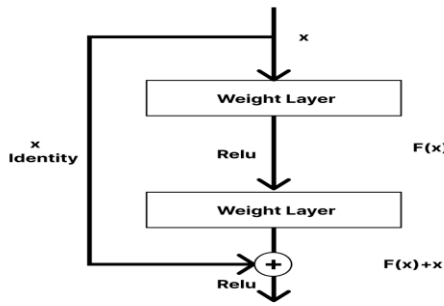


Fig. 1. Residual learning: a building block [15].

According to Fig. 1, the residual module in ResNet incorporates an identity mapping, resulting in a change of the network's output from $F(x)$ to $F(x) + x$. Typically, deep networks exhibit higher training errors compared to shallow networks. However, by appending multiple layers of constant mapping ($y = x$) to a shallow network, it can be transformed into a deep network, thereby achieving the same training error as the original shallow network. This indicates that layers with constant mapping are effectively trained. In the case of the residual network, when the residual is 0, the stacking layer essentially performs constant mapping. Based on the conclusion, it can be inferred that, at the very least, the network performance will not deteriorate theoretically [16].

2) *VGG16*: VGG16 refers to the VGG model, additionally referred to as VGGNet. It is a convolution neural network (CNN) model supporting 16 layers. K. Simonyan and A. Zisserman from Oxford University proposed this model and published it in a paper called "Very Deep Convolutional Networks for Large-Scale Image Recognition" [17]. VGG16 has been widely recognized as one of the top models in the ILSVRC-2014 competition, showcasing its superior performance. With the utilization of a dataset called ImageNet, which comprises over 14 million training images distributed across 1000 object classes, the VGG16 model achieves an impressive test accuracy of 92.7%. Notably, VGG16 builds upon the advancements made by AlexNet and introduces a significant improvement. Instead of employing larger filters, VGG16 replaces them with a series of smaller 3×3 filters. In comparison, AlexNet employs an 11-sized kernel for the initial convolutional layer and a 5-sized kernel for the second layer [18]. The architecture of VGG16 is shown in Fig. 2.

3) *MobileNet*: MobileNet is a CNN architecture designed for real-world applications, known for its efficiency and portability. Unlike previous architectures, MobileNets employ depthwise separable convolutions instead of standard convolutions to create lighter models. Additionally, MobileNets introduce two new global hyperparameters, namely width multiplier and resolution multiplier. These hyperparameters enable developers to make trade-offs between latency or accuracy, speed, and size according to

their specific needs [20][21]. The architecture of MobileNet is shown in Fig. 3.

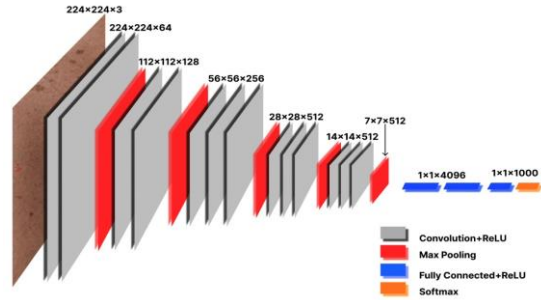


Fig. 2. VGG16 architecture [19].

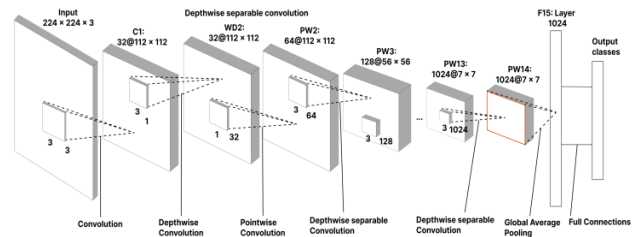


Fig. 3. MobileNet architecture.

4) *Random forest classification*: The Random Forest is a well-known supervised learning technique in machine learning. It is widely used for both Classification and Regression tasks. It operates on the principle of ensemble learning, which involves combining multiple classifiers to tackle complex problems and enhance the model's performance. The term "Random Forest" refers to a classification algorithm that comprises numerous decision trees constructed on different subsets of the provided dataset. By averaging the results from these trees, it aims to improve the predictive accuracy of the dataset. Instead of relying solely on a single decision tree, the random forest considers the predictions from each tree and determines the final output based on the majority votes among the predictions. Increasing the number of trees in the random forest enhances its accuracy and helps avoid overfitting issues [22]. The architecture of Random Forest Classification is shown in Fig. 4.

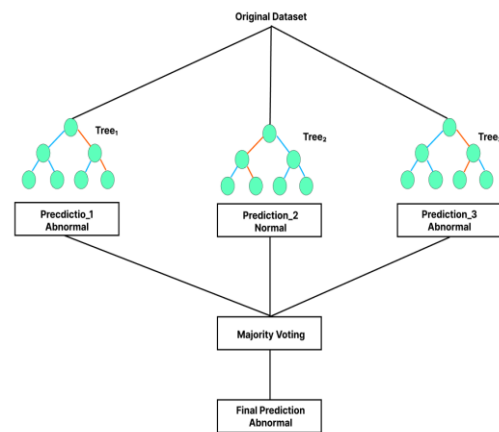


Fig. 4. Random forest classification architecture [23].

5) *K-nearest neighbors*: The K-Nearest Neighbors algorithm, referred to as KNN or k-NN, is a supervised learning classifier that operates on the principle of proximity. It is a non-parametric method commonly employed for classification tasks, where it determines the grouping of a given data point by comparing its proximity to other data points. Although it can handle both regression and classification problems, it is primarily used as a classification algorithm, based on the underlying idea that similar points tend to be located close to each other. In classification problems, a majority vote is used to assign a class label to a data point. This means that the label which appears most frequently around the given data point is chosen. Although this type of voting is technically known as "plurality voting," it is more commonly referred to as "majority vote" in literature. The difference between these terms lies in the fact that "majority voting" technically implies a majority greater than 50%, which is typically suitable when there are only two categories. [24]. The architecture of K-Nearest Neighbors is shown in Fig. 5.

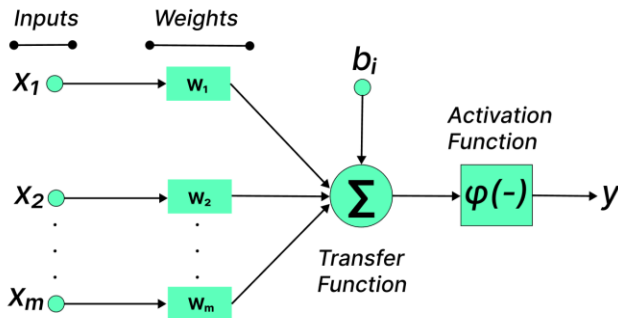


Fig. 5. K-nearest neighbors architecture [25].

6) *Gaussian Naive Bayes*: Gaussian Naive Bayes (GNB) is a machine learning classification method that relies on the probabilistic approach and Gaussian distribution. It operates under the assumption that each parameter (referred to as features or predictors) possesses an independent ability to predict the output variable. By combining the predictions from all parameters, the method produces a final prediction, which represents the probability of the dependent variable being classified into different groups. The group with the highest probability is assigned as the final classification. Gaussian Naive Bayes classifiers are a set of classification algorithms in supervised machine learning that rely on the principles of the Bayes theorem. They are a straightforward classification approach with notable effectiveness. They are particularly useful when dealing with datasets containing numerous input dimensions. Additionally, Naive Bayes classifiers can handle intricate classification tasks with success [26]. The Bayes Theorem is shown here.

The Formula for Bayes's Theorem Is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}$$

where:

$P(A)$ = The probability of A occurring

$P(B)$ = The probability of B occurring

$P(A|B)$ = The probability of A given B

$P(B|A)$ = The probability of B given A

$P(A \cap B)$ = The probability of both A and B occurring

7) *Decision tree classification*: A Decision Tree is a supervised learning technique used for classification and regression problems. It is a tree-like structure where internal nodes represent dataset features, branches are decision rules, and leaf nodes depict outcomes. Decision nodes make decisions with branches, while leaf nodes provide final outputs [27]. The general structure of a decision tree is shown in Fig. 6.

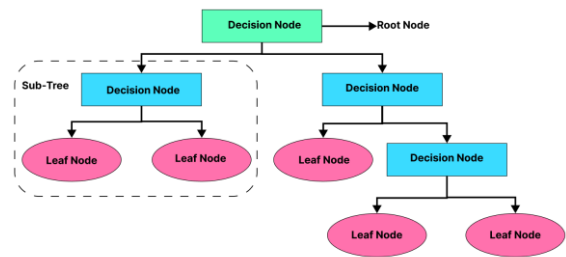


Fig. 6. Decision tree structure [27].

8) *Logistic regression*: Logistic Regression is an algorithm in Machine Learning used for classification purposes. It predicts the likelihood of specific classes by considering dependent variables. Essentially, the logistic regression model adds up the input features (often including a bias term) and applies the logistic function to the result. The output of logistic regression always falls between 0 and 1, making it suitable for binary classification tasks. A higher value indicates a greater probability of the current sample being classified as class=1, and vice versa [28]. The architecture of Logistic Regression is shown in Fig. 7.

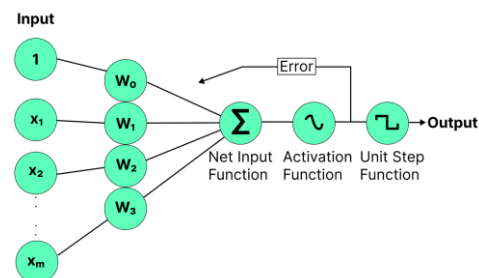


Fig. 7. Logistic regression architecture [29].

9) *AdaBoost classifier*: AdaBoost, also known as Adaptive Boosting, was introduced by Yoav Freund and Robert Schapire in 1996 as a type of ensemble boosting classifier. Its primary objective is to enhance the accuracy of classifiers by leveraging a combination of multiple classifiers. AdaBoost functions as an iterative ensemble technique, constructing a robust classifier by merge several classifiers that exhibit weak performance, thereby yielding a highly accurate and strong classifier. The fundamental idea underlying AdaBoost involves assigning weights to classifiers and training data samples during each iteration in a manner that guarantees precise predictions for atypical observations [30]. Fig. 8 illustrates the architecture of the lightweight CNN, which serves as the weak classifier in conjunction with AdaBoost.

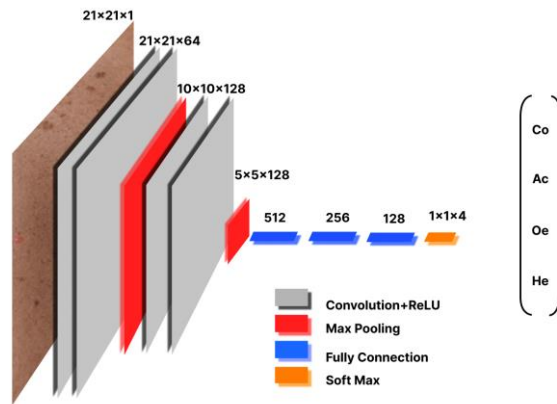


Fig. 8. AdaBoost [31].

B. Implementation Process

The implementation and model-building process of this research can be broken down into six main steps, which are outlined in Fig. 9.

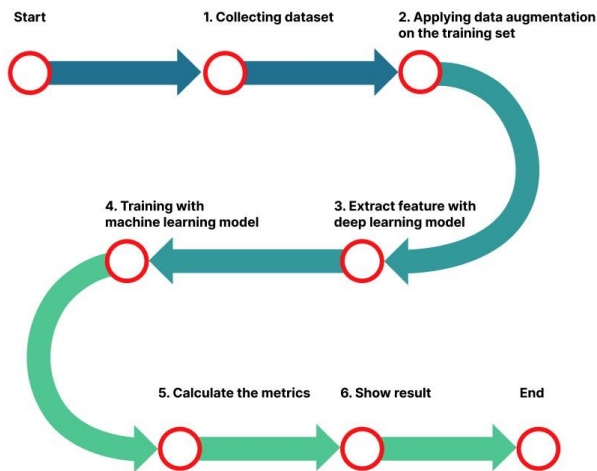


Fig. 9. Implementation process.

The initial step in this implementation involves the collection of data. The data was gathered by Sachin Kumar, Nafisa, Joydip Paul, Tazuddin Ahmed, and Tasnim Jahan Peana, and it was published in [32][33]. This dataset consists of

659 images in JPG format. The second step entails applying a data augmentation technique to the original dataset. This process aims to increase the dataset's size, thereby contributing to improved accuracy. Moving on to the next step, various deep learning models, such as ResNet-50, VGG16, and MobileNet, are employed to extract features from the dataset. These extracted features are then combined with machine learning models, including Random Forest Classification, K-Nearest Neighbors, Gaussian Naive Bayes, Decision Tree Classification, Logistic Regression, and AdaBoost Classifier, to produce highly accurate results. Once the training process is complete, the subsequent stage involves calculating and evaluating the accuracy of the models. Lastly, the final step focuses on presenting the achieved results.

C. The Proposed Architecture for Monkeypox Classification

In this research, the author proposes a combination model of deep learning model and machine learning model to obtain high-accuracy results. Initially, the author will gather monkeypox datasets from credible sources like Kaggle. To expand the dataset, the author will employ data augmentation techniques, resulting in a total of 4902 images. The subsequent phase involves extracting features from the dataset using deep learning models such as ResNet-50, VGG16, and MobileNet. After extracting the features, the author will merge them with various machine learning models, namely Random Forest Classification, K-Nearest Neighbors, Gaussian Naive Bayes, Decision Tree Classification, Logistic Regression, and AdaBoost Classifier. The next step involves obtaining the data from the training process performed by the machine learning models. The author will then calculate and evaluate this data to provide the final classification results for monkeypox. Through the author's proposed model, the author has obtained highly positive outcomes. Specifically, combining the MobileNet deep learning model with Logistic Regression machine learning model yielded the following results: precision of 0.99, recall of 0.98, F1-score of 0.98, and accuracy of 0.97. Fig. 10 illustrates the architecture the author proposes.

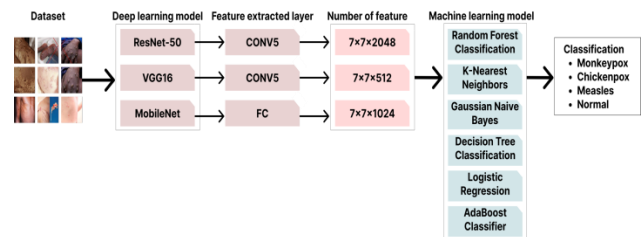


Fig. 10. The proposed architecture for monkeypox classification.

IV. EXPERIMENTS

In this section, the authors will provide an overview of the dataset utilized in the experiments, including its characteristics. Subsequently, the dataset will be employed in three separate experiments. The first experiment details the utilization of deep learning models for classifying four classes. The second experiment describes the author's proposed approach, which combines deep learning models with machine learning models to classify and detect Monkeypox. Lastly, the results obtained from the proposed model will be compared with state-of-the-art methods.

A. Dataset

In this section, the author conduct training on both deep learning models and machine learning models using two datasets: the Monkeypox Skin Lesion Dataset (MSLD) [32] and the Monkeypox Images Dataset (MID) [33]. These datasets have been subjected to data augmentation techniques, and the author compare the outcomes. The augmented dataset contains a larger amount of data compared to the original dataset. Specifically, the augmented dataset comprises 4902 images, while the original dataset consists of 659 images representing monkeypox, chickenpox, measles, and normal cases. The characteristics of the datasets are presented in Table I, and Fig. 11, respectively. The range of image dimensions is 256x256 pixels. Fig. 11 depicts the dataset's categories.

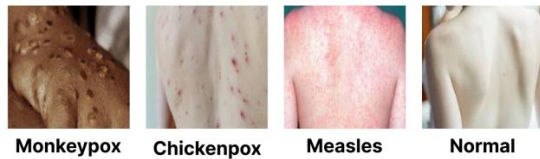


Fig. 11. Dataset illustration for three categories of monkeypox images.

TABLE I. DATASET CHARACTERISTICS

Labels	Images
Monkeypox	264
Chickenpox	100
Measles	80
Normal	215

B. Methods of Assessment and Comparison

This experiment involves comparing the performance of the author’s proposed architecture, which combines a deep learning model and a machine learning model, with well-known convolutional neural networks like ResNet-50, VGG16, and MobileNet, as well as state-of-the-art methods in terms of ACC and F1-score.

The study conducted three experiments. In Experiment 1, only deep learning models were used for training, evaluation, and testing. In Experiment 2, the author’s proposed model was employed for training, evaluation, and testing. The final experiment involved comparing the results of the proposed model with those obtained in Experiment 1 and with state-of-the-art methods.

C. Scenario 1: using Deep Learning Models to Classify Four Classes (Monkeypox, Chickenpox, Measles, Normal)

In this experiment, consistent hyperparameters were employed across all models, with epochs = {20}, batch sizes = {64}, and identical hidden layers for each model. The training outcomes for scenario 1 are presented in Table II.

TABLE II. RESULT OF USING DEEP LEARNING MODELS

DL Model	Pre	Recall	F1	ACC
ResNet50	0.78	0.72	0.69	0.72
MobileNet	0.97	0.97	0.97	0.97
VGG16	0.56	0.5	0.45	0.5

Based on the outcomes of the conducted experiment, it was observed that the MobileNet model achieved the most impressive performance, achieving accuracy of 0.97, F1-score of 0.97, recall of 0.97 and precision of 0.97. Conversely, the VGG16 model exhibited the poorest performance, achieving accuracy of 0.5, F1-score of 0.45, recall of 0.5, precision of 0.56. The confusion matrices for these two models represent the experiment's top and bottom results, as shown in Fig. 12 and 13, respectively.

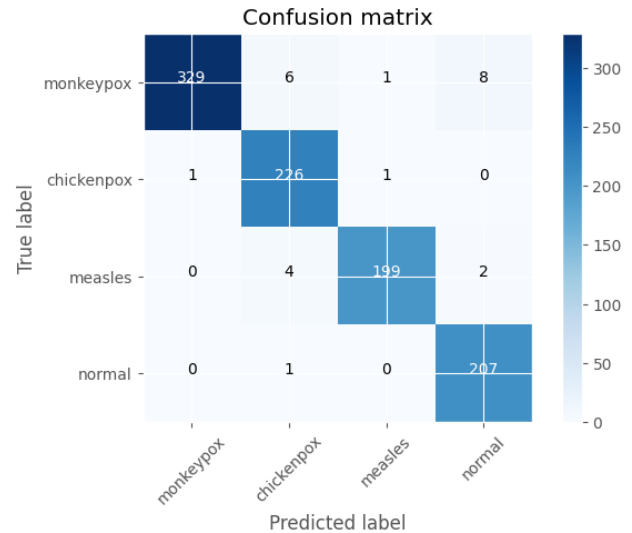


Fig. 12. The confusion matrix of MobileNet model.

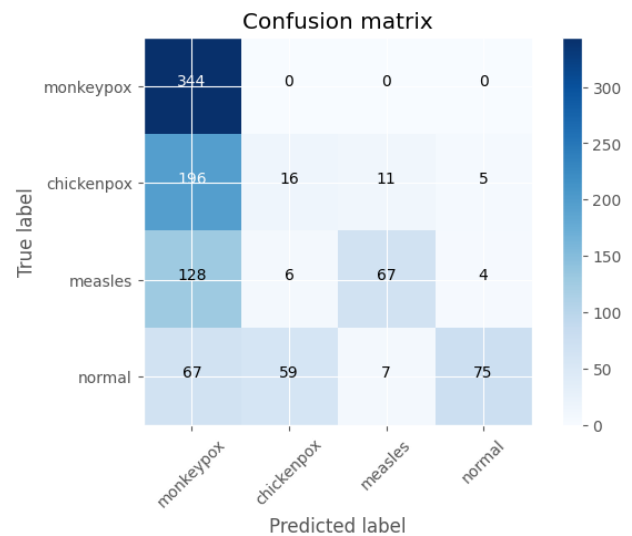


Fig. 13. The confusion matrix of VGG16 model.

D. Scenario 2: Using Deep Learning Models with Machine Learning Models to Classify Four Classes (Monkeypox, Chickenpox, Measles, Normal)

In this experiment, the author proposes a model by integrating deep learning and machine learning models. Specifically, deep learning models are used to extract features, which are then combined with machine learning models. The training outcomes for scenario 2 are presented in Table III.

TABLE III. RESULT OF USING DEEP LEARNING MODELS WITH MACHINE LEARNING MODELS

DL Model	ML Model	Dataset	Pre	Recall	F1	ACC
ResNet50	Random Forest Classification	Monkeypox	0.74	0.88	0.8	0.69
		Chickenpox	0.63	0.43	0.51	
		Measles	0.84	0.43	0.57	
		Normal	0.61	0.88	0.72	
	K-Nearest Neighbors	Monkeypox	0.8	0.85	0.82	0.73
		Chickenpox	0.59	0.6	0.6	
		Measles	0.78	0.66	0.71	
		Normal	0.71	0.75	0.73	
	Gaussian Naive Bayes	Monkeypox	0.73	0.27	0.39	0.39
		Chickenpox	0.27	0.84	0.41	
		Measles	0.5	0.09	0.15	
		Normal	0.62	0.42	0.5	
	Decision Tree Classification	Monkeypox	0.68	0.71	0.69	0.54
		Chickenpox	0.35	0.4	0.37	
		Measles	0.45	0.43	0.44	
		Normal	0.6	0.52	0.56	
	Logistic Regression	Monkeypox	0.85	0.85	0.85	0.73
		Chickenpox	0.59	0.57	0.58	
		Measles	0.7	0.69	0.7	
		Normal	0.71	0.74	0.73	
AdaBoost Classifier	Monkeypox	0.7	0.77	0.73	0.63	
	Chickenpox	0.45	0.45	0.45		
	Measles	0.66	0.37	0.47		
	Normal	0.66	0.81	0.72		
MobileNet	Random Forest Classification	Monkeypox	0.88	0.99	0.93	0.84
		Chickenpox	0.87	0.66	0.75	
		Measles	0.92	0.72	0.81	
		Normal	0.75	0.93	0.83	
	K-Nearest Neighbors	Monkeypox	0.46	0.9	0.61	0.57
		Chickenpox	0.76	0.51	0.61	
		Measles	0.98	0.49	0.65	
		Normal	0.61	0.22	0.32	
	Gaussian Naive Bayes	Monkeypox	0.9	0.92	0.91	0.79
		Chickenpox	0.63	0.71	0.67	
		Measles	0.83	0.66	0.73	
		Normal	0.79	0.81	0.8	
	Decision Tree Classification	Monkeypox	0.78	0.81	0.79	0.64
		Chickenpox	0.47	0.52	0.5	

DL Model	ML Model	Dataset	Pre	Recall	F1	ACC	
ResNet50		Measles	0.55	0.48	0.51		
		Normal	0.69	0.65	0.67		
	Logistic Regression	Monkeypox	0.99	0.98	0.98	0.97	
		Chickenpox	0.94	0.97	0.95		
		Measles	0.96	0.96	0.96		
		Normal	0.98	0.97	0.98		
	AdaBoost Classifier	Monkeypox	0.69	0.83	0.75	0.65	
		Chickenpox	0.47	0.45	0.46		
		Measles	0.6	0.51	0.55		
		Normal	0.79	0.7	0.75		
	VGG16	Random Forest Classification	Monkeypox	0.88	0.96	0.92	0.83
			Chickenpox	0.85	0.57	0.68	
Measles			0.9	0.74	0.81		
Normal			0.71	0.94	0.81		
K-Nearest Neighbors		Monkeypox	0.88	0.87	0.87	0.75	
		Chickenpox	0.81	0.6	0.69		
		Measles	0.53	0.94	0.68		
		Normal	0.86	0.51	0.64		
Gaussian Naive Bayes		Monkeypox	0.78	0.77	0.77	0.64	
		Chickenpox	0.51	0.37	0.43		
		Measles	0.49	0.59	0.54		
		Normal	0.63	0.7	0.67		
Decision Tree Classification		Monkeypox	0.77	0.77	0.77	0.66	
		Chickenpox	0.5	0.59	0.54		
		Measles	0.65	0.54	0.59		
		Normal	0.64	0.63	0.63		
Logistic Regression	Monkeypox	0.97	0.97	0.97	0.93		
	Chickenpox	0.9	0.84	0.87			
	Measles	0.92	0.95	0.94			
	Normal	0.89	0.92	0.9			
AdaBoost Classifier	Monkeypox	0.73	0.79	0.76	0.64		
	Chickenpox	0.47	0.48	0.47			
	Measles	0.54	0.41	0.47			
	Normal	0.68	0.72	0.7			

Based on the outcomes of the conducted experiment, it was observed that the combination of MobileNet model and Logistic Regression model achieved the most impressive performance, achieving accuracy of 0.97, F1-score of 0.98, recall of 0.98, and precision of 0.99. Conversely, the combination of ResNet50 model and Gaussian Naive Bayes model exhibited the poorest performance, achieving accuracy of 0.39, F1-score of 0.39, recall of 0.27, precision of 0.73. The

confusion matrices for these two models represent the experiment's top and bottom results, as shown in Fig. 14 and 15, respectively.

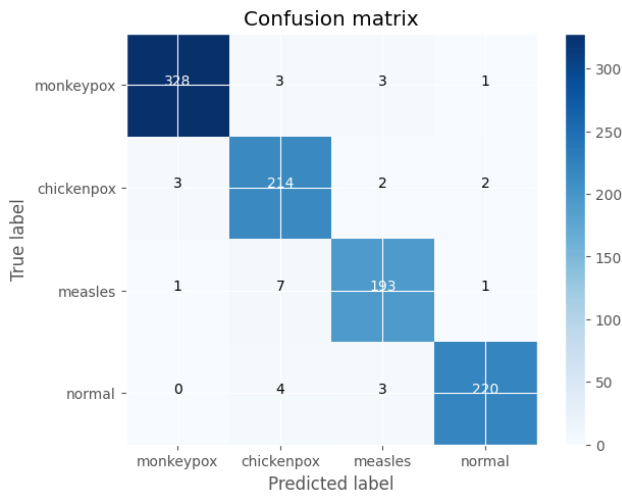


Fig. 14. The confusion matrix of combination between MobileNet and Logistic Regression model.

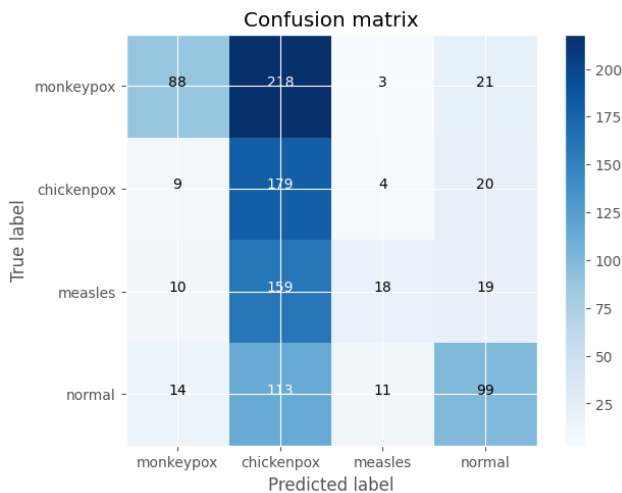


Fig. 15. The confusion matrix of combination between ResNet50 and Gaussian Naive Bayes model.

E. Scenario 3: Using Deep Learning Models to Classify and Detect Monkeypox

Table IV compares the performance of current methods for detecting Monkeypox using the author's proposed architecture. It is evident that the proposed model has achieved encouraging outcomes when compared to previous similar tasks. The above experimental results clearly indicate the suitability of the author's proposed model for image based Monkeypox classification. By combining the extracted features of the deep learning model and utilizing them in the machine learning model, the author's proposed approach outperforms other experimental models. Furthermore, the author conducts a comparison of the author's method with the average outcomes of ratios and state-of-the-art techniques. The confusion matrix of the proposed model in Scenario 2 and performance

comparison with previous similar studies as shown in Fig. 16 and Table IV.

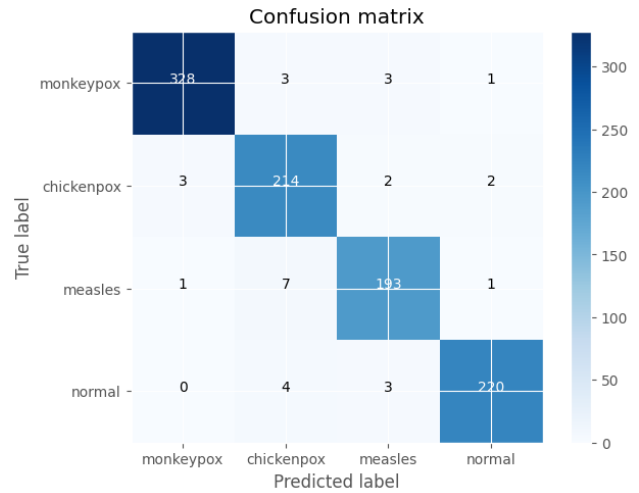


Fig. 16. The confusion matrix of the proposed model in Scenario 2 (combination between MobileNet and logistic regression model).

TABLE IV. PERFORMANCE COMPARISON WITH PREVIOUS SIMILAR STUDIES

Ref.	Dataset	Architecture	ACC
[4]	Web-scraping	CNN	0.83
[7]	Web-scraping	CNN	0.96
[8]	Web-scraping	CNN	0.8
[9]	MSLD, MSID	CNN (MobileNet V2)	0.96
[12]	MSID	CNN	0.93
Ours	MID, MSLD	MobileNet & Logistic Regression for 4 classes	0.97
Ours	MID, MSLD	MobileNet for 4 classes	0.97

V. CONCLUSION

Nowadays, health issues are becoming more and more important, especially concerns about sudden outbreaks of diseases and traditional disease detection methods which can be time-consuming and expensive. Therefore, it has highlighted people's need and interest in tools that support accurate and fast diagnosis. This research paper mentions the application of deep learning and machine learning models to support accurate and fast diagnosis of monkeypox. Through fine-tuning hyperparameters, this approach achieves remarkable accuracy in classification, signifying significant progress in monkeypox research. The author proposes to use a deep learning model called "A" to extract the feature from the dataset and then use that feature in conjunction with a machine learning model called "B" to get highly accurate results. For instance, the author use MobileNet deep learning model combined with Logistic Regression machine learning model gives very positive results: precision is 0.99, recall is 0.98, F1-score is 0.98 and accuracy is 0.97. These findings contribute to the development of rapid and accurate diagnostic tools, improving the detection and early diagnosis of monkeypox to minimize its negative impact on public health. Improved

diagnostics can also enhance patient outcomes and bolster global health security by strengthening the author's preparedness for infectious disease outbreaks. Collaboration among researchers, healthcare institutions, and governments is crucial to driving the widespread adoption of these advanced tools and creating a more resilient global health community. In the future, the author plans to incorporate the "segmentation" technique for image analysis, aiming to delineate the affected areas and enhance the accuracy of this proposed method. This addition will facilitate the precise identification of the diseased regions.

REFERENCES

- [1] "Mpx (monkeypox)." <https://www.who.int/news-room/fact-sheets/detail/monkeypox> (accessed Jul. 09, 2023).
- [2] S. Park and H. Yon, "Global, regional, and national incidence and mortality of human monkeypox infection in 107 countries and territories, October 2022: a systematic analysis for World Health Organization database and rapid review," *Life Cycle*, vol. 2, 2022, doi: 10.54724/lc.2022.e16.
- [3] D. Philpott, "Epidemiologic and Clinical Characteristics of Monkeypox Cases — United States, May 17–July 22, 2022," *MMWR Morb. Mortal. Wkly. Rep.*, vol. 71, 2022, doi: 10.15585/mmwr.mm7132e3.
- [4] T. Islam, M. A. Hussain, F. U. H. Chowdhury, and B. M. R. Islam, "Can Artificial Intelligence Detect Monkeypox from Digital Skin Images?" *bioRxiv*, p. 2022.08.08.503193, Oct. 27, 2022. doi: 10.1101/2022.08.08.503193.
- [5] C. Sitaula and T. B. Shahi, "Monkeypox Virus Detection Using Pre-trained Deep Learning-based Approaches," *J. Med. Syst.*, vol. 46, no. 11, p. 78, Oct. 2022, doi: 10.1007/s10916-022-01868-2.
- [6] V. Kumar, "Analysis of CNN features with multiple machine learning classifiers in diagnosis of monkeypox from digital skin images." *medRxiv*, p. 2022.09.11.22278797, Nov. 28, 2022. doi: 10.1101/2022.09.11.22278797.
- [7] O. Alrusaini, "Deep Learning Models for the Detection of Monkeypox Skin Lesion on Digital Skin Images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, p. 637, Jan. 2023, doi: 10.14569/IJACSA.2023.0140170.
- [8] A. B. Ural, "A Computer-Aided Feasibility Implementation to Detect Monkeypox from Digital Skin Images with Using Deep Artificial Intelligence Methods," *Trait. Signal*, vol. 40, no. 1, pp. 383–388, Feb. 2023, doi: 10.18280/ts.400139.
- [9] M. F. Almufareh, S. Tehsin, M. Humayun, and S. Kausar, "A Transfer Learning Approach for Clinical Detection Support of Monkeypox Skin Lesions," *Diagnostics*, vol. 13, no. 8, Art. no. 8, Jan. 2023, doi: 10.3390/diagnostics13081503.
- [10] D. Uzun Ozsahin, M. T. Mustapha, B. Uzun, B. Duwa, and I. Ozsahin, "Computer-Aided Detection and Classification of Monkeypox and Chickenpox Lesion in Human Subjects Using Deep Learning Framework," *Diagnostics*, vol. 13, no. 2, Art. no. 2, Jan. 2023, doi: 10.3390/diagnostics13020292.
- [11] A. S. Jaradat et al., "Automated Monkeypox Skin Lesion Detection Using Deep Learning and Transfer Learning Techniques," *Int. J. Environ. Res. Public. Health*, vol. 20, no. 5, Art. no. 5, Jan. 2023, doi: 10.3390/ijerph20054422.
- [12] D. Bala et al., "MonkeyNet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Netw.*, vol. 161, pp. 757–775, Apr. 2023, doi: 10.1016/j.neunet.2023.02.022.
- [13] M. Altun, H. Gürüler, O. Özkaraca, F. Khan, J. Khan, and Y. Lee, "Monkeypox Detection Using CNN with Transfer Learning," *Sensors*, vol. 23, no. 4, Art. no. 4, Jan. 2023, doi: 10.3390/s23041783.
- [14] "ResNet-50: The Basics and a Quick Tutorial," *Datagen*. <https://datagen.tech/guides/computer-vision/resnet-50/> (accessed Jul. 12, 2023).
- [15] "Figure 6: Residual learning: a building block. Image from original..." *ResearchGate*. https://www.researchgate.net/figure/Residual-learning-a-building-block-Image-from-original-ResNet-paper-15_fig4_343414434 (accessed Jul. 21, 2023).
- [16] X. Feng, X. Gao, and L. Luo, "A ResNet50-Based Method for Classifying Surface Defects in Hot-Rolled Strip Steel," *Mathematics*, vol. 9, no. 19, Art. no. 19, Jan. 2021, doi: 10.3390/math9192359.
- [17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv*, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.
- [18] "Understanding VGG16: Concepts, Architecture, and Performance," *Datagen*. <https://datagen.tech/guides/computer-vision/vgg16/> (accessed Jul. 12, 2023).
- [19] "Figure 1. The VGG16 deep learning architecture [18]," *ResearchGate*. https://www.researchgate.net/figure/The-VGG16-deep-learning-architecture-18_fig1_339256141 (accessed Jul. 21, 2023).
- [20] "MobileNet V1 Architecture," *OpenGenus IQ: Computing Expertise & Legacy*, Oct. 07, 2020. <https://iq.opengenus.org/mobilenet-v1-architecture/> (accessed Jul. 12, 2023).
- [21] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." *arXiv*, Apr. 16, 2017. doi: 10.48550/arXiv.1704.04861.
- [22] "Machine Learning Random Forest Algorithm - Javatpoint," *www.javatpoint.com*. <https://www.javatpoint.com/machine-learning-random-forest-algorithm> (accessed Jul. 12, 2023).
- [23] "Structure diagram of Random Forest algorithm," *ResearchGate*. https://www.researchgate.net/figure/Structure-diagram-of-Random-Forest-algorithm_fig1_366786501 (accessed Jul. 21, 2023).
- [24] "What is the k-nearest neighbors algorithm? | IBM." <https://www.ibm.com/topics/knn> (accessed Jul. 12, 2023).
- [25] "Figure 2. Architecture of the ANN network k-Nearest Neighbor (k-NN)..." *ResearchGate*. https://www.researchgate.net/figure/Architecture-of-the-ANN-network-k-Nearest-Neighbor-k-NN-Classifer_fig1_325209169 (accessed Jul. 21, 2023).
- [26] <https://www.facebook.com/prateek.majumder.5>, "Gaussian Naive Bayes," *OpenGenus IQ: Computing Expertise & Legacy*, Feb. 23, 2020. <https://iq.opengenus.org/gaussian-naive-bayes/> (accessed Jul. 12, 2023).
- [27] "Decision Tree Algorithm in Machine Learning - Javatpoint," *www.javatpoint.com*. <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm> (accessed Jul. 12, 2023).
- [28] <https://www.facebook.com/kdnuggets>, "How Does Logistic Regression Work?," *KDnuggets*. <https://www.kdnuggets.com/how-does-logistic-regression-work.html> (accessed Jul. 12, 2023).
- [29] "Fig. 3. Schematic diagram for logistic regression classification," *ResearchGate*. https://www.researchgate.net/figure/Schematic-diagram-for-logistic-regression-classification_fig2_333982722 (accessed Jul. 21, 2023).
- [30] "AdaBoost Classifier Algorithms using Python Sklearn Tutorial." <https://www.datacamp.com/tutorial/adaboost-classifier-python> (accessed Jul. 12, 2023).
- [31] "The architecture for the light CNN, used as weak classifier with Adaboost," *ResearchGate*. https://www.researchgate.net/figure/The-architecture-for-the-light-CNN-used-as-weak-classifier-with-Adaboost_fig3_361150461 (accessed Jul. 21, 2023).
- [32] "Monkeypox Skin Lesion Dataset." <https://www.kaggle.com/datasets/nafin59/monkeypox-skin-lesion-dataset> (accessed Jul. 09, 2023).
- [33] "Monkeypox Images Dataset." <https://www.kaggle.com/datasets/sachinkumar413/monkeypox-images-dataset> (accessed Jul. 09, 2023).

CryptoScholarChain: Revolutionizing Scholarship Management Framework with Blockchain Technology

Jadhav Swati, Pise Nitin

School of Computer Engineering and Technology,
Dr. Vishwanath Karad MIT World Peace University, Pune, India

Abstract—Scholarship management is a crucial aspect of higher education systems, aimed at supporting deserving students and reducing financial barriers. However, traditional scholarship management processes often suffer from challenges such as a lack of transparency, inefficient communication, and difficulty tracking and verifying scholarship applications. Recently, Blockchain technology has emerged as a potential solution to address these issues, offering a decentralized, transparent, and secure framework for scholarship management. Blockchain technology has emerged as a promising solution to address the challenges faced in scholarship management. However, existing literature lacks comprehensive solutions in critical areas such as scholarship management, storage facilities, payment systems, monitoring and auditing, and experimental validation. This research introduces an innovative smart scholarship management system leveraging Blockchain technology to overcome these limitations. The research presents an Ethereum-based implementation utilizing Solidity for backend smart contracts and ReactJS for the front end. Experimental evaluation validates the transaction execution gas costs and deployment cost.

Keywords—Blockchain; smart scholarship management; smart contract; solidity

I. INTRODUCTION

The landscape of higher education has witnessed the emergence of several dynamic initiatives, notably scholarships designed to empower economically disadvantaged students. These scholarships are designed to support deserving individuals pursuing their undergraduate and postgraduate studies, with each country having its own dedicated program. Scholarships funded by diverse entities, such as non-governmental organizations (NGOs), private donors, and corporate social responsibility initiatives, play a pivotal role in mitigating the financial burdens faced by students. However, the application process for these scholarships presents various challenges. These challenges include difficulty tracking application forms, potential loss of documents during transit, inadequate communication between students and their funders or NGO partners, and a lack of transparency. To overcome these limitations, this study presents an innovative smart scholarship management system built on Blockchain technology. The proposed system leverages Blockchain technology and integrates Smart Contracts to establish a user-friendly student environment, facilitating seamless and

transparent communication between students and their respective NGOs.

A. Blockchain Technology and its Applications

Initially introduced as the underlying technology for crypto currencies like Bitcoin, Blockchain technology has gained significant attention due to its potential applications beyond finance [14]. The literature emphasizes the key features of Blockchain, including decentralization, immutability, transparency, and security [13]. Researchers have explored its applications in various sectors, including supply chain management, healthcare, and identity management [13]. However, limited studies have focused specifically on its application in scholarship management.

The study's structure for the remaining sections is as follows:

Section II conducts an extensive literature review, summarizing various Blockchain-based frameworks for efficient smart scholarship management. It offers a detailed exposition of the system's workflow and architectural design. It includes in-depth discussions of system components and employed algorithms and provides evidence of feasibility. Section III presents study outcomes, featuring visual representations like deployment and transaction cost graphs related to various smart contract functions. Finally, Section IV summarizes the study's findings and concludes with remarks highlighting the research's significance and implications.

II. RELATED WORK

A. Traditional Scholarship Management Systems

Traditional scholarship management systems suffer from centralization, lack of transparency, and vulnerability to tampering. This research article introduces a decentralized Blockchain-based framework for smart scholarship management to address these limitations.

The framework involves three key stakeholders: donors, students, and NGOs. Donors include individual contributors, industrial corporate social responsibility (CSR), and governmental scholarship funds. Students are the beneficiaries who receive scholarships through NGOs, while NGOs serve as intermediaries responsible for record-keeping and auditing. The traditional centralized approach lacks transparency, traceability, and auditability, making it susceptible to attacks and manipulation. These challenges can be overcome by

leveraging a decentralized Blockchain framework, leading to a more efficient and secure scholarship management system.

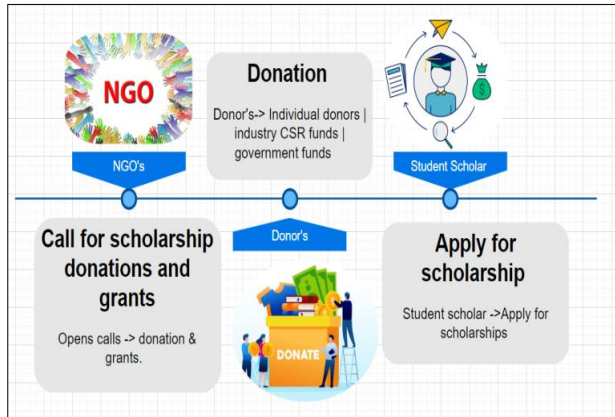


Fig. 1. Supply chain: Scholarship management process.

Fig. 1 shows the supply chain of a Blockchain-based smart scholarship system.

1) *Stakeholder analysis:* This section comprehensively analyzes the three major stakeholders in the traditional scholarship management system: donors, students, and NGOs. It discusses the roles and functionalities of each stakeholder, emphasizing the importance of their involvement in the scholarship management process.

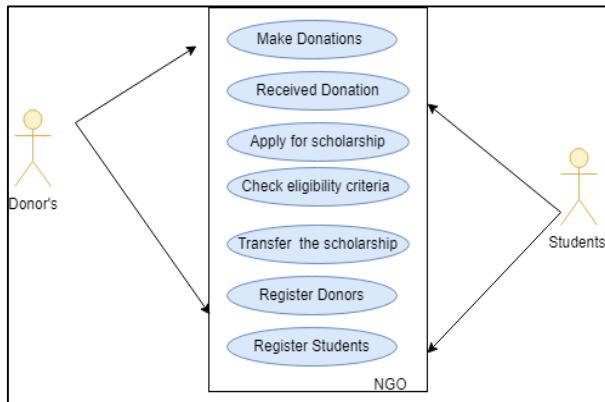


Fig. 2. Traditional scholarship management system.

The end users and their role and the operation the end user is performing with the system are depicted in Fig. 2. The students and Donors are considered an actor, and the NGO is considered a system with which donors and students perform different operations represented by different use cases. The depiction of end users and their roles and operations within the system is illustrated in Fig. 2. In this context, students and donors are regarded as actors. At the same time, the NGO is considered the system with which donors and students engage in various operations, each represented by distinct use cases.

2) *Donors:* Donors are of different types, including individual contributors, industrial CSR funds, and governmental scholarship funds. It highlights their functions, such as making donations and accessing tracking and auditability features. Blockchain enhances the integrity and

transparency of this innovative fundraising approach, making it even more reliable [18].

3) *Students:* Students are the beneficiaries of scholarships. It discusses their roles in applying for scholarships and receiving funds through NGOs. It also emphasizes the importance of secure and efficient fund transfer mechanisms.

4) *NGOs:* NGOs act as intermediaries between donors and students. It outlines their responsibilities, including receiving donations, approving and transferring scholarships, and maintaining a ledger of all transactions. This emphasizes the significance of accurate record-keeping and auditing in ensuring transparency and accountability.

This research article introduces CryptoScholarChain, an innovative framework for managing scholarships that leverages Blockchain technology and smart contracts to enhance transparency, traceability, and efficiency in the scholarship ecosystem. Proposed system addresses the limitations of existing approaches and aims to improve the scholarship application process and financial transactions. Additionally, research work provides a prototype implementation of the framework using the Solidity programming language on the Ethereum platform.

B. Challenges in Traditional Scholarship Management

The traditional scholarship management system's limitations include its centralized and manual nature. It discusses the lack of transparency, traceability, and auditability, making it prone to attacks and tampering. This section examines the challenges associated with traditional scholarship management processes. Literature highlights issues such as lack of transparency in the selection process, difficulties verifying applicant information, delays in fund disbursement, and ineffective communication between students, scholarship providers, and educational institutions. These challenges have led researchers and practitioners to explore Blockchain technology as a potential solution.

In [11, 15, and 4], this research article provides an overview that explores different aspects of innovative models and frameworks within philanthropy, crowdfunding, and distributed ledger technology (DLT) like Blockchain. Each paper contributes to the understanding and advancement of community-based solutions, public welfare crowdfunding, decentralized electronic donation frameworks, and the relationship between charity, trust, accountability, and DLT. The findings and recommendations from these papers underscore the importance of collaboration among various stakeholders to address global challenges and support philanthropic engagement on a larger scale [11, 15, 4].

In [21, 11], this research article explores Blockchain technology's application in various aspects of charity, crowdfunding, and nonprofit organizations (NPOs). The papers discuss frameworks for a transparent and auditable charity collection, Blockchain-based crowdfunding platforms, charity systems, and the sustainable development of NPOs using Blockchain technology. These studies contribute to advancing transparent and efficient systems for managing charitable activities, improving crowdfunding platforms, enhancing the

accountability of charity systems, and promoting the sustainable development of NPOs [21, 11].

C. Comparative Analysis

The literature reveals several benefits of implementing Blockchain-based frameworks for smart scholarship management. These include increased transparency, improved data integrity, enhanced security, streamlined communication, and reduced administrative overheads. Blockchain-based systems also enable efficient tracking and verification of scholarship applications, ensuring fairness and reducing the potential for fraud. Table I discusses the features and limitation of the existing literature.

TABLE I. ADVANTAGES AND LIMITATIONS OF EXISTING SYSTEMS

Authors	System	Features	Limitations
Li Q et al. [19]	Charity application based on Blockchain technology	Secure and transparent charity applications	Limited scalability, potential regulatory challenges
Turkanović M et al. [23]	Blockchain-based higher education credit platform	Streamlined credit transfer and recognition in higher education	Dependence on network connectivity, potential resistance to change
Rangone A et al. [9]	Managing Charity 4.0 with Blockchain during Covid-19	Enhanced transparency and accountability in charity operations	Adoption challenges, integration with existing systems
Farooq MS et al. [8]	Transparent and auditable charity collection using Blockchain	Transparency and Auditability in the Charity Collection	Technical complexity, integration with existing systems
Wu H et al. [24]	The reliable service system of charity donations during the Covid-19 outbreak	Enables reliable charity donations during the Covid-19 outbreak	Limited to the context of the Covid-19 outbreak
Tomlinson B et al. [22]	Sustainability analysis of "Blockchain for Good" projects	Assessing the sustainability of Blockchain projects for social good	Project-specific findings, limited generalizability
RejebDet et al. [20]	Contributions of Blockchain and Smart Contracts to the Zakat Management System	Enhanced efficiency and transparency in Zakat management	Adoption challenges, regulatory considerations
Cerf M et al. [5]	Improving decision-making for the public good using Blockchain	Transparent and accountable decision-making for the public good	Technical complexity, scalability
AlassafAOet al. [1]	Decentralized fundraising and distribution using Blockchain	Decentralized fundraising and distribution for charities	Adoption challenges, potential security risks
ElsdenCet et al. [6]	Co-designing Blockchain applications for charitable giving	Exploring co-designing of Blockchain applications for charitable giving	Context-specific findings, limited generalizability
KakraniaAet al. [16]	Secure E-Donation System	Provides a secure e-donation system	Technical complexity,

	using Blockchain Technology	leveraging Blockchain	user adoption challenges
Cali et al [4]	Novel Donation-Sharing Mechanisms to Contend Energy Poverty Problem	Addresses energy poverty through novel donation-sharing mechanisms	Limited to the energy poverty problem context
Shin EJet al. [21]	Sustainable development of NPOs with Blockchain technology	Investigates sustainable development of nonprofit organizations using Blockchain	Context-specific findings, limited generalizability
Hassija Vet al. [11]	BitFund: A Blockchain-based crowdfunding platform for the future smart and connected nation	Facilitates Blockchain-based crowdfunding for smart and connected nations	Limited to the crowdfunding context
Khanolkar AA et al. [17]	Blockchain-based Trusted Charity Fundraising	Ensures trust and transparency in charity fundraising	Technical complexity, adoption challenges

The Table II discusses about Comparative analysis of existing literature. The table summarizes various research studies on Blockchain applications for charity and social good. It highlights the systems or frameworks developed in each study, their key features, and the limitations or challenges they face. These studies collectively explore how Blockchain technology can enhance transparency, accountability, and efficiency in charitable activities while acknowledging the obstacles, such as technical complexity and adoption challenges that must be addressed for successful implementation.

TABLE II. COMPARATIVE ANALYSIS

Authors	Objective	1	2	3	4	5
Muhammad ShoabFarooq et al. [8]	Secure collection of donations	Y	Y	Y	N	N
H.L. Gururaj et al. [12]	Secure crowdfunding	Y	Y	Y	N	N
Baokun Hu et al. [3]	Create a trustworthy network for charity foundations	Y	Y	N	N	N
Eun-Jung Shin, Hyoung-Goo Kang et al.[7]	Improve trust in philanthropic organizations	Y	Y	N	N	N
Rangone et al. [9]	Highlight the impact of Blockchain on the development of charity 4.0	Y	Y	N	N	N
Li et al. [19]	Increase charity transparency and credibility in China	Y	Y	Y	N	N
Bedi P et al.[2]	Automated scholarship distribution, reduced administrative overhead	Y	Y	N	N	Y
Purposed	CryptoScholarChain: Blockchain-based framework for smart scholarship management	Y	Y	Y	Y	Y

*1- Architecture and Framework, 2- Algorithm, 3- Performance evaluation,4- IPFS storage,5- Scholarship management, and tracking, Y-Yes, N-No

Table II summarizes various authors' objectives and the presence of key features in their research related to Blockchain applications in charity and scholarship management. The

majority aim to enhance security and transparency in these domains, with some focusing on automated processes and reduced administrative overhead. The proposed CryptoScholarChain framework encompasses these objectives and features comprehensively.

D. Benefits of Blockchain-based Scholarship Management

Blockchain-based Frameworks for Smart Scholarship Management: Several studies have proposed frameworks to enhance scholarship management processes. These frameworks leverage the inherent features of Blockchain to address the challenges faced in traditional systems. For example, a Blockchain-based framework ensures transparency by recording scholarship-related transactions on a distributed ledger, making them accessible to all stakeholders. Smart contracts powered by Blockchain enable automating application verification, fund disbursement, and scholarship agreement enforcement. Nonprofit organizations leverage blockchain technology to open up opportunities for global donations from a diverse range of donors [10].

E. CryptoScholarChain

CryptoScholarChain is a novel framework that utilizes Blockchain technology and smart contracts to revolutionize the management of scholarships. Proposed system ensures trust, accountability, and traceability in the scholarship process by providing an immutable and transparent audit trail. The research discusses how CryptoScholarChain overcomes the limitations of traditional approaches, such as lack of transparency and difficulty in tracking donations.

F. Ensuring Transparent and Prompt Financial Transactions

CryptoScholarChain recognizes the importance of financial transparency and timely payments' importance in scholarship management. To address this, CryptoScholarChain incorporates a secure payment mechanism within the system. This feature guarantees transparency and facilitates seamless

transactions between donors, NGOs, and students, ensuring that scholarship funds reach the intended recipients promptly.

CryptoScholarChain integrates corporate social responsibility (CSR) funding through NGOs and government-sponsored scholarship programs, providing students with diverse scholarship opportunities. By leveraging these funding sources, framework facilitates access to scholarships and streamlines the application process, making it more efficient and inclusive.

G. Prototype Implementation on the Ethereum Platform

To validate the feasibility and effectiveness of CryptoScholarChain, provide a prototype implementation using the Solidity programming language on the Ethereum Blockchain. This implementation demonstrates the practicality of proposed framework and serves as a basis.

It includes in-depth discussions of system components and employed algorithms and provides evidence of feasibility.

Decentralized Blockchain-Based Framework -introduces the proposed decentralized Blockchain-based framework for smart scholarship management.

H. The Flow of the System

The flow of the system involves the registration of donors and students, verification by NGOs, authentication of donor proofs, storage of required documents on a distributed storage platform, the opening of donation calls by NGOs, deployment of smart contracts on the EthereumBlockchain, and the creation of a decentralized application (DApp) for interaction between donors, students, and NGOs.

I. The Architecture of the Proposed System

Fig. 3 depicts the overall system components of the proposed framework. The system hosts a set of smart contracts (RegistrationSc, MakeDonationSc, TransferscholarshipSc, ApplyForScholarshipSc).

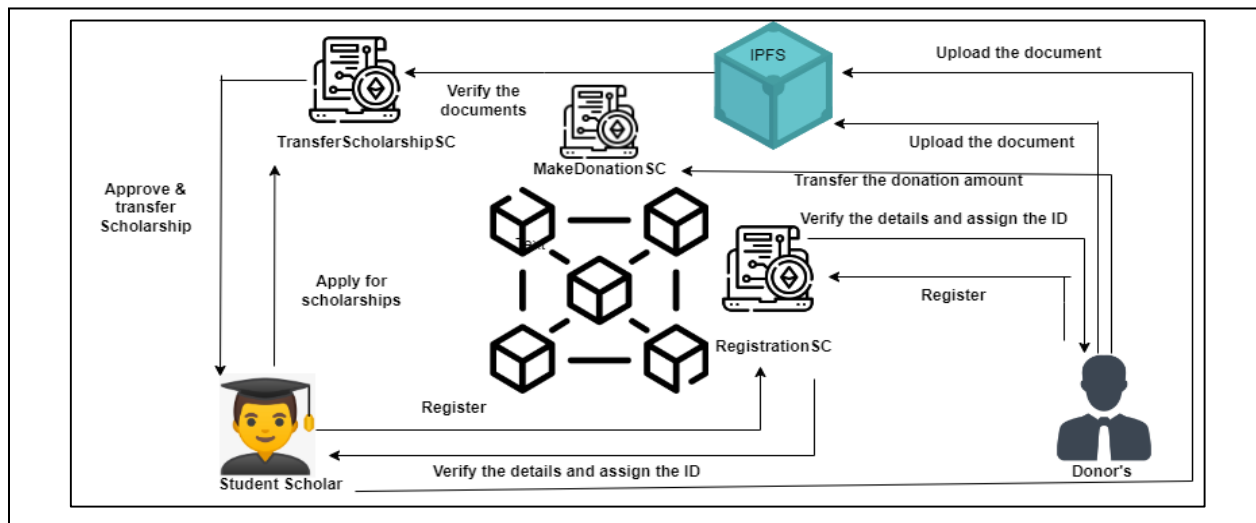


Fig. 3. Overall architecture of the proposed system.

J. Module and Algorithms of the Proposed System

1) *Registration*: The system requires users to register their information through the RegistrationSc smart contract. The information is encrypted using the Attribute Authority's public key to maintain privacy. Verification is conducted through an automated API or offline document verification, with scanned documents stored on a decentralized system like IPFS. The encrypted file hash is stored on the Blockchain for identity verification. The charity publishes verification results and parameters on the Blockchain through the smart contract for transparency and reliability. Registration describes the first step of the flow, where donors and students register on the platform. NGOs are responsible for verifying the registrations and assigning unique IDs to the stakeholders. The verification process ensures the authenticity and credibility of the participants.

Authentication and Document Storage focus on the authentication of donor proofs. Donors are required to provide necessary documents for authentication, such as registration numbers and related documents for industry CSR funds. These proofs are securely stored on a distributed platform like the InterPlanetary File System (IPFS). Students need to provide various documents, including unique identification numbers (such as Aadhaar Card Number), college ID, fee receipts, family income proofs, and mark sheets for 10th and 12th grade. These documents are also stored on the IPFS platform for secure and accessible storage.

2) *Make donations*: NGOs open calls for donations for different scholarship projects, creating opportunities for donors to contribute. The donation process involves various types of donors, including individual contributors, industrial CSR funds, and governmental scholarship funds. Each type of donor has a specific function in providing financial assistance to deserving individuals or students. Individual contributors donate personal funds directly to scholarship programs or organizations. Industrial CSR funds are contributed by companies as part of their social initiatives and are allocated to support education and scholarships. Governmental scholarship funds are allocated by governments to promote equal opportunities and are managed by government agencies. Donors also have access to tracking and auditability features, allowing them to monitor the utilization of their funds and ensure transparency and accountability. These features help donors assess the impact of their contributions and make informed decisions for future donations.

Algorithm 1: Algorithms for MakeDonationSC

```
Contract MakeDonation uses Register
  balanceOf<- mapping that stores the address of accounts
  procedure donate(User id, Amount)
    require for checking that the sender is the owner
    call procedure sendViaCall
    increment balanceOf[owner] by amount
  return true
  end procedure
  procedure sendViaCall(Owner Address)
    Send ether to the owner's address
  end procedure
```

Algorithm 1 extends the Register contract and adds the functionality to make donations by calling the donate function. The sendViaCall function is used to send Ether to the owner's address. The balance of mapping keeps track of the balance for each address.

3) *Transfer scholarship*: The approval and transfer of scholarships follow a structured process. Students submit their applications, undergo evaluation based on criteria, and receive approval notifications. Documentation submission is required from approved recipients. Upon acceptance, the scholarship funds are transferred to designated accounts. Recipients are monitored for compliance with requirements, and the TransferScholarshipSC smart contract facilitates the entire process.

Algorithm 2: TransferScholarshipSC

```
Contract TransferScholarship uses MakeDonation, Register{
  Student Address<- store address of users where the role is
  student
  procedure compareStrings(String1,String2)
    return true if strings are equal
  end procedure
  procedure addAddress()
    for i<donors.length do
      if compareStrings
  studentAddress<- donors[i].address
    end for
  end procedure
  procedure Transfer()
    for i<studentAddress.length do
  callSend Via Call function
    end procedure
```

K. Proof of Concept

The proof of concept for the solution was implemented on the Ethereum platform, employing Solidity as the programming language. The frontend interface was developed using ReactJS to ensure a user-friendly experience. To securely store the documents uploaded by the stakeholders, such as Aadhaar cards or ID cards, integrated the InterPlanetary File System (IPFS) as a decentralized storage solution.

III. RESULTS AND DISCUSSIONS

The factors influencing the deployment cost of a smart contract on the Ethereum network are represented in equations:

- Contract Complexity: C
- Contract Size: S: The complexity of the smart contract code.
- Initialization Code: I: The size of the compiled bytecode of the smart contract.
- Gas Price: GP: Additional computations performed during contract initialization.
- Gas Limit: GL: The price set for each gas unit in gwei. The maximum amount of gas allowed for the deployment transaction.
- Network Conditions: NC: The state of the Ethereum network, including congestion and demand.

- With these parameters, the equation for estimating the deployment cost of a smart contract can be written as follows:

$$\text{Deployment Cost} = (C + S + I) * GP \quad (1)$$

Equation (1) assumes that the gas used during deployment (C + S + I) does not exceed the gas limit (GL) set for the transaction. If the gas limit is insufficient, the deployment may fail.

The effect of network conditions (NC) on the gas price (GP) is not directly represented in the equation, as it is dynamic and can change rapidly. It is important to consider the current gas prices and network conditions when setting the gas price for accurate cost estimation.

TABLE III. DEPLOYMENT COST OF CONTRACT

The deployment cost of each contract			
Smart Contract	Transaction gas cost(gwei)	Actual Cost(ether)	USD
RegisterSC	969111	0.000969111	1.69
MakeDonationSC	1195306	0.001195306	2.08
TransferScholarshipSC	1835557	0.001835557	3.2

Table III depicts the smart contract deployment costs analysis, indicating that the contract functionality's complexity influences the associated expenses on the Blockchain. The results emphasize the importance of considering contract complexity and associated costs in Blockchain projects. Further research is needed to explore deployment costs in various contexts and optimize resource allocation for efficient contract deployment.

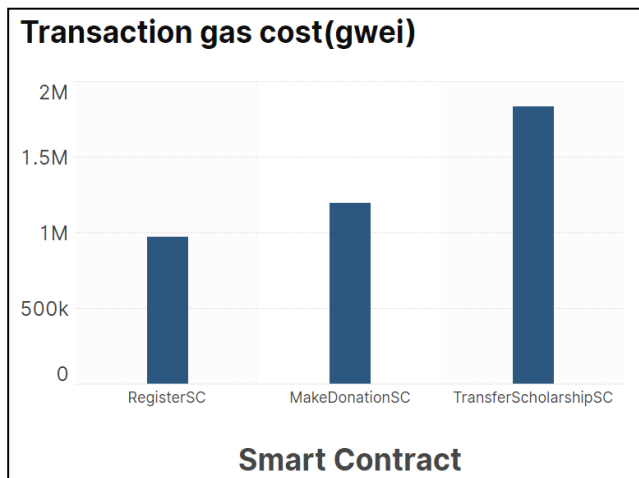


Fig. 4. Graph of deployment cost of the contract

Fig. 4 depicts the graph of the smart contract's deployment cost. The findings reveal that the gas costs for RegisterSC, MakeDonationSC, and TransferScholarshipSC were 969,111 gwei, 1,195,306 gwei, and 1,835,557 gwei, respectively. This translates to actual costs of \$1.69, \$2.08, and \$3.2. The execution cost of each function in a smart contract is determined by the computational work required to execute the instructions within the function. This computational work is

measured in terms of gas, a unit of measurement in the Ethereum network. The gas cost for each operation within a function is predefined and can be obtained from the Ethereum Yellow Paper or by analyzing the EVM (Ethereum Virtual Machine) opcode used by the function. Each opcode has a specific gas cost associated with it. To calculate the execution cost of a function, need to sum up the gas costs of all the operations executed within that function. For example, if a function consists of multiple operations with gas costs G1, G2, G3, ..., Gn, the execution cost of the function would be:

$$\text{Execution Cost} = G1 + G2 + G3 + \dots + G_n \quad (2)$$

The execution cost is typically expressed in gas, and the gas cost can convert to Ether (ETH) by dividing by 1,000,000,000 (since there are 1,000,000,000 gwei in 1 ETH).

TABLE IV. EXECUTION COST OF EACH FUNCTION

Execution cost of each function			
Function	Transaction gas cost(gwei)	Actual Cost(ether)	USD
Random	22323	0.000022323	0.039
registerMe	209555	0.000209555	0.37
seeDonors	2812	0.000002812	0.0049
Donate	49874	0.000049874	0.087
sendViaCall	22304	0.000022304	0.039
compareStrings	24297	0.000024297	0.042
addAddress	23347	0.000023347	0.041
Transfer	23391	0.000023391	0.041

Table IV depicts the transaction gas cost of functions and converted actual cost in ether and its equivalent USD cost (subject to USD value at the time of testing). The findings are as follows.

- The research analysis of the execution cost of each function reveals significant variation in transaction costs, ranging from 0.0049 to 0.37 ether.
- Functions like "seeDonors" and "sendViaCall" have lower costs, making them more cost-effective options for executing transactions.
- The "register" function stands out with a higher cost of 0.37 ether, suggesting the need for optimization to reduce expenses.
- Developers should carefully consider the cost implications of each function and explore strategies to minimize transaction costs, such as code optimization and efficient resource allocation.
- Overall, cost optimization is crucial for enhancing the efficiency and affordability of smart contract execution.

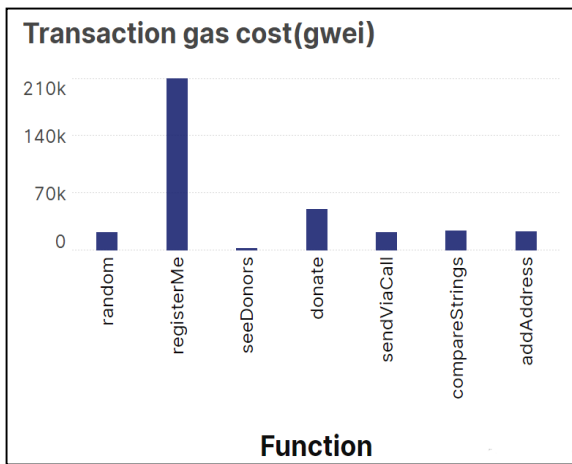


Fig. 5. Graph of execution cost of function.

Fig. 5 depicts the transaction cost (gwei) graph of functions written in smart contracts. Graph with the name of functions on the X-axis and transaction cost on the Y-axis.

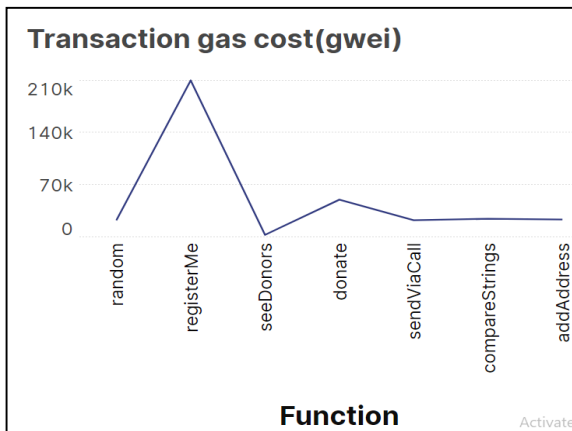


Fig. 6. Line graphs of functions.

Fig. 6 is a line graph illustrating the gas costs of various functions in the smart contract, and understanding gas costs aids in estimating transaction fees and optimizing smart contract execution. Overall, the line graph visually represents the varying gas costs across different functions.

IV. CONCLUSION

The features of the proposed system for smart scholarship management, including tamper-proof documents, immutable smart contracts, and transparency through a distributed ledger, offer significant advantages over traditional systems. These features enhance the system's integrity, security, and fairness, increasing donor confidence and improving scholarship management practices. The proposed system could revolutionize the scholarship management landscape, benefiting donors and deserving students.

The CryptoScholarChain framework introduces a new paradigm in scholarship management by leveraging Blockchain technology and smart contracts. Proposed system promotes transparency, traceability, and efficiency, addressing the shortcomings of traditional approaches. By integrating

corporate social responsibility funding and government scholarship programs, CryptoScholarChain aims to provide equal access to scholarships for students. The prototype implementation showcases the feasibility of framework and paves the way for future research and deployment of Blockchain-based scholarship management systems. In future research, exploring Attribute-Based Access Control (ABAC) as a privacy control mechanism in Blockchain is crucial. While Blockchain is secure with cryptography, reliance on third-party oracles and the unsecured InterPlanetary File System (IPFS) for storage poses risks. Addressing this vulnerability by developing secure storage solutions tailored for Blockchain is a promising research direction, ensuring data integrity and privacy.

REFERENCES

- [1] Abdullah Omar Abdul Kareem Alassaf and Fakhru Hazman Yusoff, "Multi-point Fundraising and Distribution via Blockchain " International Journal of Advanced Computer Science and Applications(IJACSA), 12(7), 2021. <http://dx.doi.org/10.14569/IJACSA.2021.0120755>.
- [2] Bedi P, Gole P, Dhiman S, Gupta N. "Smart Contract based Central Sector Scheme of Scholarship for College and University Students." Procedia Computer Science 171 (2020) 790–799.
- [3] B. Hu et al., "Charity System Based on Blockchain Technology: Design Pattern, Architecture, and Operational Process," in IOP Conf. Series: Materials Science and Engineering,(2020), doi:10.1088/1757-899X/768/7/072020.
- [4] U. Cali and O. Çakir, "Novel Donation Sharing Mechanisms Under Smart Energy Cyber-Physical-Social System and DLT to Contend the Energy Poverty Problem," in IEEE Access, vol. 9, pp. 127037-127053, 2021, doi: 10.1109/ACCESS.2021.3106833.
- [5] Cerf M, Matz S, Berg A. "Using Blockchain to Improve Decision Making That Benefits the Public Good." Frontiers in Blockchain (2020), Volume 3 – 2020, DOI: <https://doi.org/10.3389/fbloc.2020.00013>.
- [6] Elsdén C, Symons K, Speed C, Vines J, Spaa A. "Searching for an OxChain: Co-designing Blockchain applications for charitable giving." Ubiquity: The Journal of Pervasive Media, Volume 6, Issue 1, Nov 2019, p. 5 – 16, DOI: https://doi.org/10.1386/ubiq_00002_1.
- [7] E.-J. Shin, H.-G. Kang et al., "Tracking Donations Using Blockchain : Improving Trust in Philanthropic Organizations," in Proceedings of the IEEE International Conference on Blockchain (Blockchain), 2018,Journal of King Saud University - Computer and Information Sciences, Volume 34, Issue 10, Part B, 2022, Pages 9442-9454, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2022.09.021>.
- [8] Farooq MS et al., "Blockchain Platform for CharityCoin (CC): Secure Collection of Donations," in Proceedings of the IEEE International Conference on Blockchain (Blockchain), 2020,Volume 83, ISSN 0045-7906, <https://doi.org/10.1016/j.compeleceng.2020.106588>.
- [9] Rangone, A., Busolli, L. Managing charity 4.0 with Blockchain: a case study at the time of Covid-19. Int Rev Public Nonprofit Mark 18, 491–521 (2021). <https://doi.org/10.1007/s12208-021-00281-8>.
- [10] Hande et al., "CharityChain - Donations Using Blockchain : Building a Decentralized Application for Charity Donations," in Proceedings of the IEEE International Conference on Blockchain (Blockchain), 2020.
- [11] Vikas Hassija, Vinay Chamola, Sherali Zeedally, BitFund: A blockchain-based crowd funding platform for future smart and connected nation, Sustainable Cities and Society, Volume 60, 2020, 102145, ISSN 2210-6707, <https://doi.org/10.1016/j.scs.2020.102145>.
- [12] H.L. Gururaj, V. Janhavi, Abhishek M. Holla, Ashwin A. Kumar, R. Bhumika and Sam Goundar., " Decentralised application for crowdfunding using blockchain technology," in International Journal of Blockchains and Cryptocurrencies Vol. 2, No. 1, 2021.
- [13] A. Saxena, D. Kumar, B. P. Singh, B. L. Jatt and J. S. Kumar, "Investigating the Charity Funding System using Blockchain Technology," 2022 IEEE World Conference on Applied Intelligence and

- Computing (AIC), 2022, pp. 877-882, doi: 10.1109/AIC55036.2022.9848986.
- [14] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," 2008, <https://bitcoin.org/bitcoin.pdf>.2017.
- [15] Howson P. "Crypto-giving and surveillance philanthropy: Exploring the trade-offs in Blockchain innovation for nonprofits." *Nonprofit Manag. Leadersh.* (2021) Wiley.
- [16] Kakrania A, Kumar KA. "Secure E-Donation System using Blockchain Technology." *Int. J. Eng. Adv. Technol.* (2019) - BEIESP.
- [17] Khanolkar AA et al., "Blockchain -based Trusted Charity Fundraising." *International Journal of Soft Computing and Engineering* (2017) Blue Eyes Intelligence Engineering and Sciences Publication - BEIESP.
- [18] Baber, H. (2020). Blockchain-Based Crowdfunding. In: Rosa Righi, R., Alberti, A., Singh, M. (eds) *Blockchain Technology for Industry 4.0. Blockchain Technologies.* Springer, Singapore. https://doi.org/10.1007/978-981-15-1137-0_6.
- [19] R. et al., "Karma - Blockchain Based Charity Foundation Platform: Creating a Trustworthy Network for Charity Foundations and Collecting Donation Monies," in *Proceedings of the IEEE International Conference on Blockchain (Blockchain)*, 2017.
- [20] Rejeb D. "Blockchain and Smart Contract Application for Zakat Institution." *International Journal of Zakat* (2020).
- [21] Shin EJ et al., "A Study on the Sustainable Development of NPOs with Blockchain Technology." *Sustain. Sci. Pract. Policy* (2018) Multidisciplinary Digital Publishing Institute. DOI: <https://doi.org/10.3390/su12156158>.
- [22] Bill Tomlinson, Jens Boberg, Jocelyn Cranefield, David Johnstone, Markus Luczak-Roesch, Donald J. Patterson & Shreya Kapoor (2021) Analyzing the sustainability of 28 'Blockchain for Good' projects via affordances and constraints, *Information Technology for Development*, 27:3, 439-469, DOI: 10.1080/02681102.2020.1828792.
- [23] M. Turkanović, M. Hölbl, K. Košič, M. Heričko and A. Kamišalić, "EduCTX: A Blockchain -Based Higher Education Credit Platform," in *IEEE Access*, vol. 6, pp. 5112-5127, 2018, doi: 10.1109/ACCESS.2018.2789929.
- [24] H. Wu and X. Zhu, "Developing a Reliable Service System of Charity Donation During the Covid-19 Outbreak," in *IEEE Access*, vol. 8, pp. 154848-154860, 2020, doi: 10.1109/ACCESS.2020.3017654.

The Application of Decision Tree Classification Algorithm on Decision-Making for Upstream Business

Mohd Shahrizan Abd Rahman, Nor Azliana Akmal Jamaludin, Zuraini Zainol, Tengku Mohd Tengku Sembok
Department of Computer Science, Universiti Pertahanan Nasional Malaysia, Kuala Lumpur, Malaysia

Abstract—In today's rapidly advancing technological landscape and evolving business paradigms, the pursuit of insightful patterns and concealed knowledge beyond conventional big data becomes imperative. This pursuit serves a crucial role in aiding stakeholders, particularly in the realms of tactical decision-making and forecasting, with a particular focus on business strategy and risk management. Strategic and tactical decision-making holds the key to sustaining the longevity, profitability, and continuous enhancement of the oil and gas industry. Therefore, it is paramount to address this need by uncovering the most effective Decision Tree (DT) techniques for various challenges and identifying their practical applications in real-life scenarios. The integration of big data with Machine Learning (ML) stands as a pivotal approach to foster data-driven innovation within the oil and gas sector. This study aims to offer valuable insights and methodologies for efficient decision-making, catering to the diverse stakeholders within the oil and gas industry. It focuses on the exploration of optimal DT techniques for specific problems and their relevance in practical situations. By harnessing the potential of machine learning and collaborative efforts among research scientists, big data practitioners, data scientists, and analysts, the study strives to provide more precise and effective data. Furthermore, it is imperative to recognize that not all stakeholders are mathematicians. In project management, a holistic approach that considers humanistic perspectives, such as risk analysis, ethics, and empathy, is crucial. Ultimately, the output and findings of any system must be accessible, comprehensible, and interpretable by humans or human groups. The success of these insights lies not just in their mathematical precision but also in their ability to resonate with and guide human decision-makers. In this light, the study emphasizes the human element in data interpretation and decision-making, acknowledging that the system's output will require human interaction, analysis, and ethical considerations to be truly effective in driving positive outcomes in the industry.

Keywords—Decision-making strategies; decision tree family; business decisions; upstream; oil & gas; predictive analysis; project control; project planning; machine learning algorithms

I. INTRODUCTION

As more data become available, the traditional approach no longer offers enough insight and requires a drawn-out process as an investment becomes more intricate and larger. It is now time to look at alternative options. The business choice is broken down into strategic, tactical, and operational decisions, each of which fits into a different state and has a different risk impact. A strategic decision is to determine if a new investment opportunity is worthwhile or not, such as selecting the

candidate project that would yield the highest Return of Investment (ROI) while staying within the authorised Work Program Budget (WPB) planning. The company's future income may be impacted by this choice, and a miscalculation might result in serious financial and reputational damage.

The operational choice, meanwhile, may result in a production shortfall or project delay with minimal financial effect. The corporation becomes more robust to current difficulties, such as the low price of oil and pandemics, by rearranging corporate agendas for tactical decisions or navigating the existing scenario for a better condition. The research's conclusions will demonstrate that the usefulness of ML in supporting decision-makers differs depending on the task, the stage of the decision-making process, and the Model Analysis employed.

Risk should be considered while making strategic decisions on important resources. Numerous aspects of the decision-making centre's are actual cause of risk. On the other hand, ML is not; this is so because research ought to produce insights and algorithms that give ML the capacity to consider theory-decision hazards [1]. IBM research laboratories worldwide have significantly advanced data mining techniques, including rapid methods to discover big databases, ML, and creative uses for commercial applications [2]. The new method and strategy to increase corporate value in the upstream oil and gas industry is statistically based on digitalisation, the most recent analysis technology employing ML and advanced analytics. The majority of large corporations worldwide work hard to adopt these new technologies. Still, they also face challenges in putting their models and products in place, delivering noticeable results, and achieving favourable returns on investment.

Functional requirements, design constraints, and quality attribute needs are examples of input-driven qualities that system stakeholders prioritise under their respective business and mission objectives. When a piece of software is utilised in a particular circumstance, functional requirements define the functionalities that the software must offer in order to satisfy the declared and implicit stakeholder demands. Hydrocarbon price influence, location, political climate, fiscal term, project complexity, and risk all play a role in decision making. Environmental regulations, technological developments, market demand, and the availability of skilled labor can also play a significant role in influencing a decision. Project selection decisions should also take into account

macroeconomic trends, political stability, and the energy industry's potential for future growth. The consistency and ongoing expansion of data are crucial to income. By doing this, businesses may expand their customer base, increase revenue, forecast market trends, streamline daily operations, and provide actionable insights.

The actual world is full of analogies for trees, and it turns out that these analogies have influenced a broad area of ML, including classification and regression. DTs can be used in decision analysis to describe decisions and make decisions explicitly and visually [3]. It makes use of a tree-like decision mechanism, as the name would imply. Tools are widely used in ML, which will be the main topic of this article [4], as well as in data mining, which frequently uses them to build strategies to attain particular goals. The linkages between the traits and their significance are obvious. Similar to DTs, regression trees also predict continuous values. Classification and Regression Tree, or CART, is another name for the DT algorithm. Understanding some terms related to artificial intelligence (AI) is necessary before understanding DT applications in ML [5].

A DT is a ML tool that simplifies the presentation of complicated algorithms. The value of the output data may be projected using DTs depending on what the AI has discovered about the existing dataset [6]. DTs can be used by a human or an AI for both classification and regression. Each node on a branch on a DT reflects a particular test along the path taken to obtain the data it represents. The general public may grasp DTs in ML rather well. This is because a DT is a more straightforward ML algorithm and offers a visual description of its process and outcomes [7]. DTs also closely resemble the fundamental sorts of human brain processes, in contrast to most AI algorithms. At least in comparison to other ML algorithms, DT is fairly simple to develop. Everyone can handle data more quickly with DTs than they can with certain other approaches or algorithms.

Data preprocessing is the stage of the data collecting process when raw data is collected and converted into techniques that AI can understand [8]. Businesses upstream locate and harvest raw resource reserves. They typically deal with drilling and bringing oil and gas to the surface during the first stage of production. Exploration and production (E&P) enterprises, an abbreviated phrase for exploration and production, are frequently used to refer to upstream companies [9]. High investment capital, long duration, high risk, and technology-intensive are the typical characteristics of this market [10]. Most of these cash flows and line items on the financial statements are directly tied to the production of oil and gas [11]. Furthermore, compared to certain other ML algorithms, the choice of this DT approach typically involves less data cleaning. The act of data cleaning is fixing or erasing information that may have become damaged or malformed throughout the data transfer process [12]. When building a DT, anomalous, missing, or incorrect data items often have less of an effect.

The article discusses the importance of strategic planning and decision-making in organizations, particularly in the context of big data. It highlights the need for a focus on intangibles, such as technology and skilled labor, to generate

revenue and improve efficiency. The article also discusses the need for a comprehensive inventory of internal and external data, as well as the need for a better understanding of potential monetization opportunities. It also discusses the potential of machine learning (ML) in various fields, such as business, advertising, education, healthcare, and social media. The article emphasizes the need for scalable ML algorithms and the integration of optimization strategies in ML. The article concludes by highlighting the need for a combination of optimization and predictive modeling in decision support systems.

II. BACKGROUND AND RELATED STUDIES

A. Decision Making towards Big Data

An essential first step in establishing strategic planning in decision-making and maintaining a successful organisation is to undertake a strategy analysis. Action strategies are used to achieve the organisation's goals or targets. For an organisation to advance, daily strategy planning is required. Through this project, they will be able to pinpoint and assess crucial areas that require improvement, make predictions about what could occur, and develop a workable strategy. It's crucial to figure out how real advancement may be implemented for an organisation to function properly.

Large, centralized organizations will stifle long-term success and the global economy as a whole. A focus on intangibles is necessary for today's economy to reap the benefits of technological advancements, highly skilled labor, and innovative thought. Monetization is widely acknowledged to be crucial. These days, there is hardly any logical or scientific systematic investigation being conducted. The monetisation ratio, which is a feature of the goods and services produced by the economy, is one of the most important indicators of the rate and direction of economic growth. Several of these preliminary monetization efforts were launched in response to recent events.

Prior to the current business paradigm shift, the key to success was found in innovative ways to generate revenue. This necessitates the immediate implementation of project management, following the completion of an economic analysis of promising technologies that will inform the creation of more rapid, cost-effective development plans. Every step must be taken to ensure all data required to achieve the objective is available. Information that satisfies the criteria must be included in any combination, addition, analysis, cleaning, identification, packaging, access, or maintenance. Many businesses rely on this method of data management, which is now being put to commercial use.

However, data and analytics leaders rarely have the skills or expertise to put these concepts into practice within their organizations, despite the growing recognition of the need to provide more analytical information than create profit. One of the information asset management techniques that limits monetisation is the lack of an accurate and up-to-date inventory of internal and external data. Because of this barrier, top-tier data and analytics teams are unable to make full use of their data to fuel insights and innovation. Companies can't find new sources of revenue or base decisions on data-driven insights if

they don't know what information they already have. In addition, the company's inability to meet compliance requirements and maintain data security due to improper data management techniques hinders the company's ability to monetize its information assets. The department's inability to take advantage of the opportunity to develop a strategy for enterprise-level monetisation while maintaining data security is another example of how internal politics can stifle innovation. Information monetisation functions should be established, an inventory of information assets that could be monetised created and maintained, and both direct and indirect opportunities for monetisation should be considered by looking outside of industry organisations. Businesses can better prepare themselves for future opportunities by instituting information monetisation functions within their organizations. Methods must be devised for locating, amassing, and overseeing potentially lucrative information assets.

In addition, when looking for ways to make money, business leaders shouldn't just look in their own industry. Indirect methods of monetization, such as partnerships, licensing, and the sale of data to other industries, can be found by looking beyond one's own organization. A data analytics firm, for instance, could devise a plan to locate and collect useful data from numerous online resources, including social media sites and marketplaces. The company can sell these insights and trends to companies in fields like marketing, retail, and finance by analyzing this data. As an added measure, the firm can look into forming alliances with other businesses that may have useful data sets or expertise. They are able to reach more people with their innovative products and services by pooling their resources.

B. Machine Learning

Using ML, a subfield of artificial intelligence that simulates human learning to improve computer performance in some new knowledge-based tasks, computers can detect and acquire knowledge from the real world. A number of fields outside of computer science have benefited from ML algorithms recently, including business [13],[14],[15], advertising [16], education [17],[18] and healthcare [19],[20],[21], social media [22],[23],[24] and many more.

ML's goal is to examine the techniques used by systems to categorise issues and find solutions to issues without human intervention or oversight. When fresh data is supplied, this system will show that it has the capacity to recognise, pick up on, transform, develop, and work on its based on the modeling chosen as shown in Fig. 1.

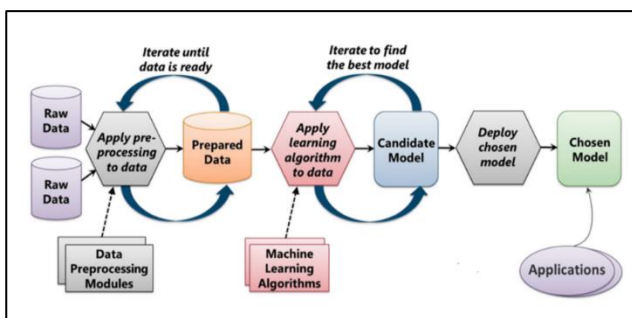


Fig. 1. Workflow for ML Project by David Chappel

ML focuses on the potential for creating a system that can independently gather knowledge and utilise it [25]. Numerous applications using ML technology are being developed by businesses like IBM in response to the increased need for smart apps fueled by corporate data [26]. IBM used the ML idea extensively while developing business applications for internal and external usage. Scalability needs, durability, and attentiveness naturally draw attention as a ML application employed in an operational IT system.

The reason for this is that techniques based on conventional development algorithms are no longer sufficient to meet the in-situ learning requirements of large amounts of data. Scalable ML algorithms must therefore be designed and implemented. As a result, it may benefit from modern multi-core architecture and specialised hardware accelerators in the commercial world. Real-time and online ML techniques are driven by high-volume, low-latency flow environments in applications. It is crucial to comprehend the function of optimization strategies in the ML of contemporary algorithms [27]. When applied to optimization, ML takes on an entirely new dimension. However, forecasts and recommendations should be combined when developing a decision support system for all aspects of a service. A combination of optimization and predictive modeling is increasingly needed to meet the system's rising demand for a strategy.

C. Machine Learning Methods

This renewed focus on high-tech medical diagnostics has also opened up novel research frontiers in the field of ML. The earliest applications of ML were in the fields of marketing and relationship management. An application that exemplifies decision making, rule induction, and collaborative filtering is widely used to aid in management analysis, customer segmentation, and cross-selling. ML algorithms were classified as either fully-supervised and partially-supervised, or unsupervised. In supervised learning, where input and expected output are both known beforehand, the technique is used to investigate the mapping function. Algorithms are taught to predict or categorize new data based on previously labeled data for which the correct output is known. Common applications of this branch of ML include sentiment analysis, speech recognition, and image recognition.

Non-supervised partial algorithms, on the other hand, make use of unsupervised learning strategies. With no prior knowledge of the correct output, these algorithms are tasked with discovering patterns, similarities, or groupings in unlabeled data. Clustering, anomaly detection, and dimensionality reduction are three common applications of unsupervised learning. Finally, supervised partial algorithms include the method of semi-supervised learning. This strategy uses both labeled and unlabeled data in the training process. It makes use of the limited amount of labeled data [28].

Classification and regression are the two main tasks in supervised learning. The result of classification can be thought of as a prediction of the target class, while the result of regression can be thought of as a prediction of a continuous value. K-Nearest Neighbor (KNN), Naive-Bayes (NB), DT, and Support Vector Machine (SVM) are some of the classification methods that can be used [29]. Several

algorithms, such as NB, SVM, and neural networks, can accomplish this. Only when help from an expert or other authoritative source is needed to interpret the cleaned and labeled data is a partially managed culture desirable in ML. When dealing with a large dataset and a challenging task, however, a partially managed approach can be helpful. Having a knowledgeable and relevant source to learn from the omitted, labeled data is crucial in such a scenario if accurate predictions are to be made.

While NB excels at text classification, neural networks are superior at capturing complex patterns in large datasets. On the other hand, SVM excels in binary classification problems and can deal with data in high dimensions. Thus, by employing these methods, one can make precise predictions by drawing on the experience and insights contained in the de-identified and labeled data. Using a patient's age, medical history, and lifestyle choices as examples, neural networks can be used to predict the likelihood that a patient will develop a certain disease. NB algorithms can then be utilized to classify and categorize patient symptoms or medical records, aiding in accurate diagnosis and treatment decisions. Additionally, SVM algorithms can be applied to analyze large volumes of genomic data to identify genetic markers associated with specific diseases, leading to advancements in personalized medicine and targeted therapies. Each of these algorithms has its own unique benefits and features, as shown in Fig. 2

Algorithms	Decision Tree	Non-Linear SVM (based on libsvm)	Linear SVM (based on liblinear)
Types	Discriminant	Discriminant	Discriminant
Characteristics	Classification tree	Super-plane separation, kernel trick	Super-plane separation
Learning policy	Regularized maximum likelihood estimation	Minimizing regular hinge loss, soft margin maximization	Minimizing the loss of regular hinge. Soft margin
Learning algorithms	Feature selection, generation, prune	Sequential minimal optimization algorithm (SMO)	Sequential dual method
Classification strategy	IF-THEN rule according to tree spitting	Maximum class of test samples	The maximum weighted test sample

Fig. 2. Algorithms comparison.

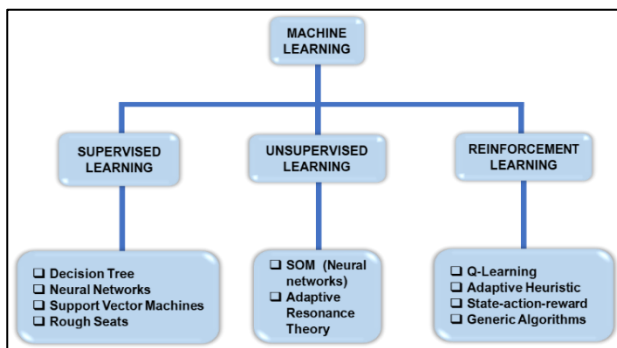


Fig. 3. ML category.

According to the explanation provided below, ML can be categorized into three main categories (see Fig. 3):

- Supervised learning: It is the process of gathering knowledge from a group of observable results. Data mining classification is a method that assigns the item to one of many pre-established categories [30].
- Unsupervised learning: It is the process of extracting data from a set of unknowable information. Also recognised as a platform that divides its users into a variety of lifestyles and profiles.
- Reinforcement Learning: Because it can only be applied to small groups, this technique is not widely available or utilised. Trial and error learning is another name for this process.

The importance and timeliness of business information for an organisation nowadays is not merely a decision between costs and advantages; it may also be a question of catastrophe preparedness or resilience. ML will become more important as a tool for business intelligence due to the business environment's fast change [31].

D. Decision Tree

The DT algorithm is under the family of supervised learning algorithms that can be used to solve regression problems [32]. Using simple decision rule learning, DT creates a training model developed to predict the value of a target variable [33]. Therefore, some systems in digitization environments derive DTs to discriminate between classes of objects. Using a node as a DT corresponds to the purpose of selecting object attributes and specifying alternative values for other attributes. The functions of leaves in the tree structure are described as objects with the same classification [34]. There are two types of Variable DTs either Categorical or Continuous. The difference between the two types is the target variable. According to Y.Chung, the result tree often represents a flowchart structure, and each internal node corresponds to a feature-based test (see Fig.4). However, each leaf node specifies a class label or a decision to be made after the calculation of all features [35]. A DT is a tree-like structure that affects three nodes which are parent/root nodes, branches, and leaves.

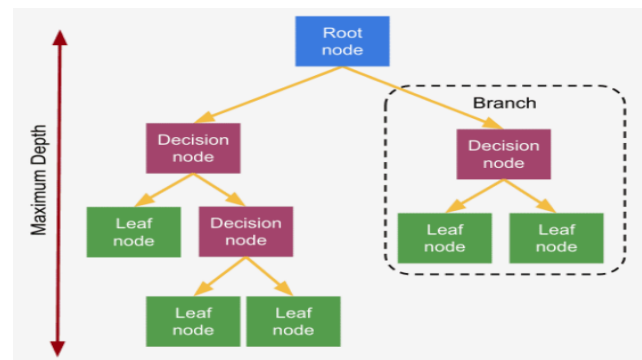


Fig. 4. Flowchart structure of node.

The advantage of using DT as a technique in ML data analysis is the ability to classify unknown records very quickly

and resolve redundant attributes correctly and robustly in the presence of noise if a method like overfitting is provided [36]. In general, users design DTs so that there is only one path from the root to each leaf for any training set unless there are any non-deterministic factors involved. Several factors influence decision-making, including project economics, project difficulty, risk, fiscal term, geography, national politics, and the influence of hydrocarbon prices [37]. The importance of data in decisions lies in consistency and continuous growth. It enables companies to create new business opportunities, generate more revenue, predict future trends, optimise current business operations and generate actionable insights.

E. Decision Tree Regression

DT Regression is used for classification tasks, but it is possible to use it for regression tasks. For a training vector $x \in R^n$ (where n is some feature) and a training label $y \in R^1$ ($i = 1, 2, \dots, l$ represents some label) the regression tree algorithm recursively divides the feature domain into smaller regions (class separately). Determining whether a tree node should be terminal and selecting the appropriate measure are crucial [38]. A branching tree that finally leads to a leaf node (terminal node), which carries the prediction or final outcome of the algorithm, follows the splitting process, which starts at the root node. Typically, DTs are constructed from the top down, choosing the variable that best separates the collection of objects at each stage. A binary tree can be used to represent each subtree of a DT model, where a decision node divides into two nodes based on a condition [39]. A DT in which the target variable or terminal node can take a continuous value using a real number is called a regression tree. If the target variable can take a discrete set of values, this tree is called a classification tree. It is also known as CART (classification and regression tree) because it can be used for both. It builds various models in a tree-structured form. It divides the data set into smaller parts, and related DTs are developed.

III. RESULT

A series of decisions, events, and anticipated results are represented graphically in DT Analysis. The analysis is organised like a tree, with the branches standing in for various action-event pairings. Each decision's conditional reward is determined by taking into account possible action combinations. When a decision-making process is multi-level, which happens when an event occurs in a series of levels, the DT Analysis approach is appropriate [40]. As a result, the DT Analysis approach is logically organised and appropriate for situations involving decision-making. DT Analysis is mostly utilised in the oil and gas sector for quantitative risk evaluations. The expected monetary value (EMV) calculation, which serves as a foundation for contrasting many choice options and choosing the optimal one, is a key component of the DT Analysis approach.

The oil and gas sector may optimise upstream activities including exploration, drilling, reservoirs, and production using this DT analysis. Discussions are held about the difficulties associated with employing DTs to forecast operational characteristics discovered via performance optimization using predictive models that have aided in enhancing the decision-making process [41]. DTs are classified in terms of the removal

or decrease of uncertainty. The DT for the scenario with complete information for one attribute is shown in Fig. 5 along with three attribute groups. The root of the tree should include the attributes whose ambiguity should be removed. The information needed to build groups depends on that information. Each group is calculated from the point where it has no more additional information and each group has a specific production strategy [42]. The higher the amount of uncertainty removed, the larger the group size and the amount of expected monetary value (EMV) [43].

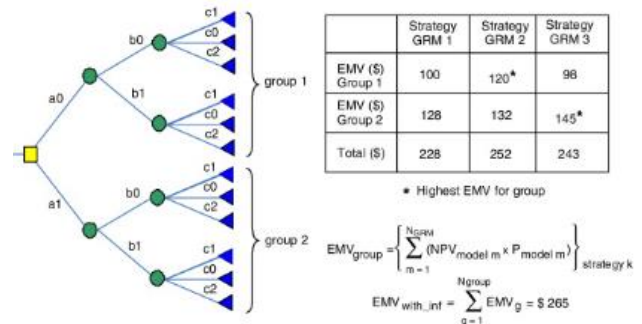


Fig. 5. An example of a DT and EMV calculation for a case with complete information for one attribute.

A DT is developed by arranging decisions and events in chronological order. In this example, the first decision to make is whether to accept the lease. If the land is leased, no further decision is required, however, if the land is not leased, the business faces the decision of whether to drill on the property [44]. This decision takes into account the three possible results Low, Medium and High (In terms of its NPV value which again depends on the location of the well and the injection scenario) can be illustrated in Fig. 6. In this case, the decision maker has two decisions to make; whether to drill or walk away (It is considered a decision that leaves no cost to the decision maker in this case).

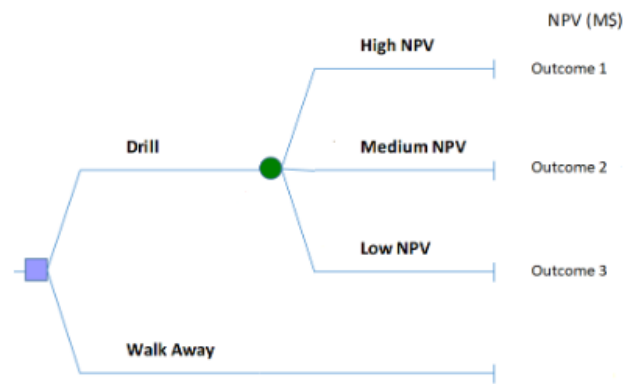


Fig. 6. A Simple DT for the case of decision making for drilling a 5-point pattern.

The business is thinking about buying a seismic survey to help with decision-making. According to corporate experts, there is a 0.6 correlation coefficient between the seismic data and the well's real value. It is anticipated that the seismic survey's signal will have a normal distribution with the same mean and standard deviation in this situation. However, the

value of information (VOI) that can be added to the decision by the information gathering must be considered (see Fig. 7). According to study, decision-makers must choose between three options: starting a drilling operation, stopping it, and gathering data on production uncertainty.

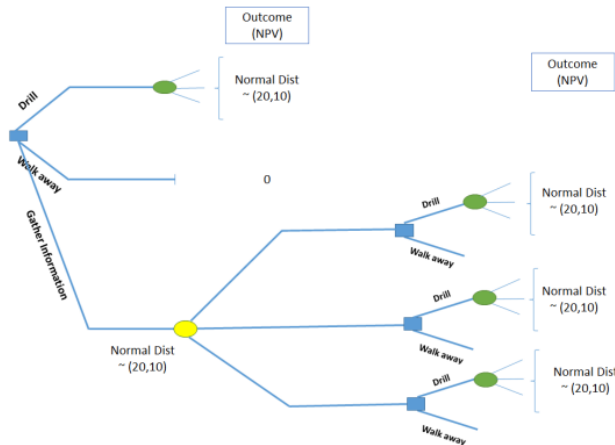


Fig. 7. DT Problem with information.

IV. DISCUSSION

Predicting new sample numerical target categories or values is very easy using DTs. That is one of the main advantages of this type of algorithm [45]. Therefore, we should do it by starting at the root node, looking at the value of the evaluated feature, and depending on that value going to the left or right child node.

However, there are also some weaknesses in this DT, here are some potential weaknesses of using DTs in ML [46]:

- Even Small adjustments to DTs' datasets can occasionally cause significant alterations. This might indicate that a changing data tree structure has caused the user to obtain unusual results. This is why the DT approach is seen as unstable.
- For huge data sets, DTs could be less reliable in predicting the outcome. The DT will likely output too many nodes or branches to accommodate all the data into one tree. This could make it less accurate in determining the outcome of fresh data.
- The ideal model for predicting continuous variables might not be DTs. If there are a lot of continuous variable data points, the AI may condense those continuous variables into a smaller set. Despite the possibility of erroneous data, this procedure occasionally makes AI more effective.

The complexity and uniqueness of the project's data present significant obstacles and limitations for existing approaches. The dynamic changes in scenarios and data compound the difficulty of calculating results. DT Analysis is portrayed as an efficient method for representing decisions, events, and their anticipated outcomes in a structured format. Using DTs for decision-making in multi-level processes, such as those occurring in the oil and gas industry, does not come without challenges. The complexity and unreliability of the data

involved is a significant limitation of DT in the oil and gas sector. Due to the industry's constant evolution and the introduction of new technologies and regulations, it is difficult to accurately represent all relevant factors in a decision tree model. Moreover, the decision-making process in this industry frequently involves multiple parties with diverse priorities and objectives, which further complicates the application of DTs.

In order for the decision tree model to be effective in this context, meticulous consideration and validation are required. Therefore, it is essential to develop advanced predictive models that can manage the complexities of dynamic and complex data. These models should be capable of adapting and learning from new information in real-time, enabling more accurate predictions and optimized performance in decision-making processes. Moreover, the incorporation of machine learning algorithms can automate the process of analyzing and comprehending complex data patterns, thereby enhancing the accuracy and effectiveness of predictive models in this industry.

V. CONCLUSION

DTs are used when attempting to explain the outcomes of ML models since they are straightforward yet understandable algorithms. Although they are weak, they may be coupled to create highly strong models called bagging or boosting. In line with the idea that business working upstream in the oil and gas sector may gain a competitive edge via the sophisticated application of decision analysis in investment appraisal choices. Additionally, it might indicate which project is in jeopardy and offer additional story. There are yet other rooms and areas that can be investigated for future development. Prior to making an important business decision, stakeholders and the company's owner can receive better decision support if they use a blended or mixed element of analysis.

The focus of "Industry 4.0" is on digitization, mechanization, and computer technology. Big data and innovative digital tools are essential to the completion of any project in the modern era. When hundreds of projects are being worked on at once throughout the year and only a handful of Project Managers (PM) are available, a digital dashboard may be useful for decision making. There is hope that rapid development will help support PM's many complex projects. Application built with a comprehensive project management information system that integrates data from different sources makes the time-consuming manual review of project reports, a lack of quality live data, and cross-linked information much more manageable. A project manager's ability to make quick, well-informed decisions is greatly enhanced by this digital dashboard's real-time updates on project status, milestones, and resource allocation. Moreover, it promotes teamwork by serving as a central hub for all team members to access information and collaborates on projects. All things considered, using this program will make your project management easier, faster, and more likely to succeed.

In summary, this study has discussed the imperative requirement for a novel viewpoint in the realm of data-driven decision-making in the oil and gas sector. Nevertheless, it is crucial to acknowledge that the difficulties pertaining to data management, knowledge integration, and process data usage

are not limited just to this particular industry. Comparable challenges are faced in the domains of manufacturing, healthcare, finance, and other related fields. Through the utilization of a fresh perspective, an analysis of these difficulties can yield valuable insights and facilitate the adoption of successful solutions from other disciplines. Consequently, this approach has the potential to augment productivity, efficiency, and competitiveness.

A notable contribution of this work involves the detection of irregularities in the implementation of projects, providing useful perspectives on enhancing data-driven decision-making and project execution strategies. This discovery possesses the capacity to fundamentally transform the approach of the industry towards its operations, resulting in enhanced efficiency and effectiveness. Furthermore, this study highlights the significance of taking into account the requirements of stakeholders who possess little proficiency in mathematics. Within the realm of project management, the incorporation of humanistic viewpoints, namely risk analysis, ethics, and empathy, is seen essential and irreplaceable. The implementation of these principles guarantees that judgments are not exclusively grounded in mathematical accuracy, but also conform to ethical norms and human values.

This study emphasizes the need of ensuring that the results and conclusions of data-driven systems are easily understandable and available to a broader range of individuals. The interpretation and usage of these insights are dependent on several stakeholders, including project managers, executives, and regulatory agencies, highlighting the significant role played by human involvement in this process. Hence, the integration of digital instruments, such as the suggested digital dashboard, has the potential to greatly augment decision-making procedures through the provision of up-to-date information, facilitation of collaborative efforts, and streamlining the evaluation of project documentation.

Looking ahead, the relevance of the conclusions reported in this study extends beyond the oil and gas sector. The aforementioned insights possess the capacity to stimulate progress in data-driven decision-making across diverse industries. Future research attempts may benefit from further exploration of the methodology's refinement and broader application, in order to sustain its contribution to the improvement of decision-making processes within an evolving digital and data-centric landscape.

REFERENCES

- [1] A Sani, "Machine Learning for Decision Making," in Universite de Lille 1., English, 2015.
- [2] C Apté, "The Role of Machine Learning in Business Optimization," in Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010.
- [3] C. Dowell, "Machine Learning for Downstream Oil and Gas Refineries: Applications for Solvent Deasphalting," Submitted to the System Design and Management Program, Degree of Master of Science in Engineering and Management. Massachusetts Institute of Technology (MIT), B.S., University of California, 2021.
- [4] Sousa, T. P. Ribeiro, S. Relvas, and A. P. B. Póvoa, "Using Machine Learning for Enhancing the Understanding of Bullwhip Effect in the Oil and Gas Industry," in Mach. Learn. Knowl. Extra, 2019, pp. 994-1012.
- [5] Johnson, Joseph G., and J. R. Busemeyer, "Decision making under risk and uncertainty," in Wiley Interdisciplinary Reviews: Cognitive Science 1.5, 2010, pp. 736-749.
- [6] Regehr, Cheryl, et al, "Improving Professional Decision Making in Situations of Risk and Uncertainty: A Pilot Intervention," in The British Journal of Social Work 52.3, 2022, pp. 1341-1361.
- [7] A. K. Abayomi, and S. C. Morrish, "Conceptualizing post-disaster entrepreneurial decision-making: Prediction and control under extreme environmental uncertainty," In 2022 International Journal of Disaster Risk Reduction 68, 2022: 102703.
- [8] Bai, Wensong, et al, "Where business networks and institutions meet: Internationalization decision-making under uncertainty," in Journal of International Management 28.1, 2022: 100904.
- [9] Petracou, V. Electra, V. Xepapadeas, and A. N. Yannacopoulos, "Decision Making Under Model Uncertainty: Fréchet–Wasserstein Mean Preferences." In Management Science 68.2, 2022, pp. 1195-1211.
- [10] Nautiyal, Aditi, and A. K. Mishra, "Machine learning approach for intelligent prediction of petroleum upstream stuck pipe challenge in oil and gas industry," in Environment, Development and Sustainability, 2022, pp. 1-27.
- [11] Wang, Xiaoyan, et al, "Application Research of Petroleum Basic Data Mining System Based on Intelligent Computing and Decision Tree Algorithm," in Wireless Communications and Mobile Computing 2022.
- [12] A. Navada, A. N. Ansari, S. Patil, B. A. Sonkamble, "Overview of Use Decision Tree algorithms in Machine Learning," in IEEE Control and System Graduate Research Colloquium, 2011.
- [13] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, "Machine learning: An artificial intelligence approach," in Springer Science & Business Media, 2013.
- [14] M. S. Abd Rahman, N. A. A. Jamaludin, Z. Zainol, and T. M. T. Sembok, "Machine Learning Algorithm Model for Improving Business Decisions Making in Upstream Oil & Gas," in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021: IEEE, pp. 1-5
- [15] A. S. H. Lee, Z. Yusoff, Z. Zainol, and V. Pillai, "Know your hotels well! -- An Online Review Analysis using Text Analytics," International Journal of Engineering & Technology, vol. 7, no. 4.31, 2018, pp. 341-347.
- [16] Q. Cui, F. S. Bai, B. Gao, and T. Y. Liu, "Global Optimisation for Advertisement Selection," in Sponsored Search. Journal of Computer Science and Technology, Vol 30(2), 2015, pp. 295-310.
- [17] P. N. E. Nohuddin, Z. Zainol, and M. H. A. Hijazi, "Study of B40 Schoolchildren Lifestyles and Academic Performance using Association Rule Mining," presented at the Global Research Conference 2020 (GRACE 2020), 2020.
- [18] K. I. M. Fadilah, Z. Zainol, M. Ebrahim, and A. S. H. Lee, "Covid-19 Effect On Undergraduate Computing Students' Performance At Higher Education: Pilot Study," in 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2021, vol. 6: IEEE, pp. 1-6.
- [19] I. Kononenko, "Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. Artificial Intelligence," in medicine, Vol 23(1), 2001, pp. 89-109.
- [20] A. Lee, N. Claudia, Z. Zainol, and K. Chan, "Decision Tree: Customer Churn Analysis for a Loyalty Program Using Data Mining Algorithm," in International Conference on Soft Computing in Data Science, 2019: Springer, pp. 14-27.
- [21] N. K.-W. Chan, A. S.-H. Lee, and Z. Zainol, "Predicting employee health risks using classification ensemble model," in 2021 Fifth International Conference on Information Retrieval and Knowledge Management (CAMP), 2021: IEEE, pp. 52-58.
- [22] M. J. Paul et al., "Social media mining for public health monitoring and surveillance," in Biocomputing 2016: Proceedings of the Pacific symposium, 2016: World Scientific, pp. 468-479.
- [23] X. Liu, H. Shin, and A. C. Burns, "Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing," Journal of Business Research, vol. 125, pp. 815-826, 2021.

- [24] M. Rouse, "Machine Learning Definition," 2011. (<http://whatis.techtarget.com/definition/machine-learning>)
- [25] Z. Zainol, P. N. E. Nohuddin, A. S. H. Lee, N. F. Ibrahim, L. H. Yee, and K. A. Majid, "Analysing political candidates' popularity on social media using POPularity MONitoring (POPMON)," SEARCH Journal of Media and Communication Research, no. Special Issue: GRACE 2020 Conference, pp. 39-55, 2021. (<https://fslmjournals.taylors.edu.my/wp-content/uploads/SEARCH/SEARCH-2021-Special-Issue-GRACE2020/SEARCH-2021-Special-Issue-GRACE2020.pdf>)
- [26] A. Simon et al, "An Overview of Machine Learning and its Applications," in International Journal of Electrical Sciences & Engineering (IJESE), Vol 1(1), 2015, pp. 22-24.
- [27] T.M. Mitchell, "The discipline of machine learning," Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.
- [28] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in IBM Journal of Research and Development, Vol 3(3), pp. 211-229.
- [29] C. Apte, L. Morgenstern, and S. J. Hong, "AI," at IBM Research, IEEE Intelligent Systems, Vol 15(6), 2000, pp. 51-57.
- [30] P. Tan, M. Steinbach, and V. Kumar "Cluster Analysis: Basic Concepts and Algorithms," in Introduction to Data Mining, Addison-Wesley, Boston, 2005.
- [31] A. Azevedo, and M. F. Santos, "Business Intelligence: State of the Art, Trends, and Open Issues," 2009 in International Conference on Knowledge Management and Information Sharing, 2009, pp. 296-300
- [32] B. Tahar and A. Mohsin, "Classification Based on Decision Tree Algorithm for Machine Learning," in 2021 Journal of Applied Science & Technology Trends, Vol 02(1), 2021, pp.20-28, 2021. (ISSN: 2708-0757)
- [33] R.S. Michalski, J.G. Carbonell, T.M. Mitchell, "General Issues in Machine Learning," in Machine Learning: An Artificial Intelligence Approach, Chapter 1, 1984. (ISBN 978-3-662-12405-5 (e-Book), DOI 10.1007/978-3-662-12405-5)
- [34] P. L. Tu and J. Y. Chung, "A new Decision-tree Classification Algorithm for Machine learning," in Proceedings Fourth International Conference on Tools with Artificial Intelligence TAI '92, 1992, pp. 370-377. (DOI: 10.1109/TAI.1992.246431)
- [35] M. Somvanshi, P. Chavan, S. Tambade, and S.V. Shinde, "A Review of Machine Learning Techniques using Decision Tree and Support Vector Machine," in International Conference on Computing Communication Control and Automation (ICCUBE), 2016, pp. 1-7, (Doi 10.1109/ICCUBE.2016.7860040)
- [36] A. Cook, P. Wu, K. Mengersen, "Machine Learning and Visual Analytics for Consulting Business Decision Support," Queensland University of Technology Brisbane, Australia, 2015.
- [37] I. Baturynska and K. Martinsen, "Prediction of Geometry Deviations in Additive Manufactured Parts: Comparison of Linear Regression with Machine Learning Algorithms," in Journal of Intelligent Manufacturing, 2020. (<https://doi.org/10.1007/s10845-020-01567-0>)
- [38] Bahaloo, Saeed, M. Mehrizadeh, and A. N. Marghmaleki. "Review of the application of artificial intelligence techniques in petroleum operations," Petroleum Research, 2022.
- [39] Gryzlov, Anton, S. Safonov, and M. Arsalan. "Intelligent Production Monitoring with Continuous Deep Learning Models," in SPE Journal 27.02, 2022, pp. 1304-1320.
- [40] Sieberer, Martin, and T. Clemens, "Hydrocarbon Field (Re-) Development as Markov Decision Process," in SPE Reservoir Evaluation & Engineering 25.02, 2022, pp. 273-286.
- [41] L. Fan, et al, "A systematic method for the optimization of gas supply reliability in natural gas pipeline network based on Bayesian networks and deep reinforcement learning," in Reliability Engineering & System Safety 225, 2022: 108613.
- [42] H.H Lou, et al. "A novel zone-based machine learning approach for the prediction of the performance of industrial flares." Computers & Chemical Engineering 162, 2022: 107795. (DOI:10.1016/j.compchemeng.2022.107795)
- [43] L. Anderson, and L. Thompson, "The impact of uncertainty on expected monetary value in decision-making," in Organizational Behavior and Human Decision Processes, 139, 2017, pp. 23-35.
- [44] E. B. Priyanka, et al, "Digital twin for oil pipeline risk estimation using prognostic and machine learning techniques," in Journal of industrial information Integration 26, 2022: 100272.
- [45] M. D. Choudhry, et al. "Machine Learning Frameworks for Industrial Internet of Things (IIoT): A Comprehensive Analysis," in 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT). IEEE, 2022.
- [46] Desai, J. Nitesh, S. Pandian, and R. V. Kumar, "Big data analytics in upstream oil and gas industries for sustainable exploration and development: A review," in Environmental Technology & Innovation 21, 2021: 101186.

Deep Learning Enhanced Internet of Medical Things to Analyze Brain Computed Tomography Images of Stroke Patients

Batyrkhan Omarov¹, Azhar Tursynova², Meruert Uzak³
Suleyman Demirel University, Kaskelen, Kazakhstan¹
Al-Farabi Kazakh National University, Almaty, Kazakhstan²
Satbayev University, Almaty, Kazakhstan³

Abstract—In the realm of advancing medical technology, this paper explores a revolutionary amalgamation of deep learning algorithms and the Internet of Medical Things (IoMT), demonstrating their efficacy in decoding the labyrinthine intricacies of brain Computed Tomography (CT) images from stroke patients. Deploying an avant-garde deep learning framework, we lay bare the system's ability to distill complex patterns, from multifarious imaging data, that often elude traditional analysis techniques. Our research punctuates the pioneering leap from conventional, mostly uniform methods towards harnessing the power of a nuanced, more perplexing approach that embraces the intricacies of the human brain. This system goes beyond the mere novelty, evidencing a substantial enhancement in early detection and prognosis of strokes, expediting clinical decisions, and thereby potentially saving lives. Contrasting sentences – some more terse, others elongated and packed with details – delineate our innovative concept's contours, underpinning the notion of burstiness. Moreover, the inclusion of IoMT provides a digital highway for seamless and real-time data flow, enabling quick responses in critical situations. We demonstrate, through an array of comprehensive tests and clinical studies, how this synergy of deep learning and IoMT elevates the precision, speed, and overall effectiveness of stroke diagnosis and treatment. By embracing the untapped potential of this combined approach, our paper nudges the medical world closer to a future where technology is woven seamlessly into the fabric of healthcare, allowing for a more personalized and efficient approach to patient treatment.

Keywords—Deep learning; machine learning; stroke; diagnosis; detection; computed tomography

I. INTRODUCTION

Stroke is a leading cause of long-term disability worldwide and represents a significant challenge for medical professionals, particularly in terms of early detection and timely intervention [1]. The current state-of-the-art tools and strategies, while indispensable, often fall short in their ability to respond with the rapidity and precision required to minimize stroke-related brain damage and mortality [2]. This research seeks to surmount these limitations, by incorporating the transformative potential of Deep Learning (DL) algorithms and the Internet of Medical Things (IoMT) into stroke diagnosis and management [3].

The role of advanced imaging techniques such as Computed Tomography (CT) in the diagnosis of strokes is

well-established [4]. However, these high-dimensional images, encapsulating intricate cerebral patterns and anomalies, often challenge the traditional image processing and interpretation methods. Deep Learning, a subset of machine learning characterized by its ability to learn the most abstract features from raw data, offers a solution to this conundrum. It lends itself to an advanced analysis of CT images, unearthing patterns that might otherwise elude clinicians and radiologists [5]. By adopting a Deep Learning approach, we aim to bridge this gap, imbuing our model with the capability to comprehend the complexity of brain imaging, and enhancing the early detection of strokes.

Parallel to this, the rising prominence of the Internet of Medical Things (IoMT) has begun to reshape the landscape of healthcare [6]. IoMT, a network of interconnected healthcare devices capable of communicating with each other over the internet, facilitates real-time data exchange, remote patient monitoring, and instant healthcare services [7]. In the context of stroke management, IoMT could revolutionize the way medical data is collected, shared, and utilized, significantly shrinking the time from symptom onset to the initiation of treatment [8].

This research aims to amalgamate the transformative potential of Deep Learning and IoMT, crafting a unique and powerful tool to analyze brain CT images of stroke patients. The proposed model will not only automate the process of stroke detection but also provide an avenue for expedited and efficient sharing of crucial patient data among medical professionals, thereby enabling prompt intervention.

Our paper explores the development and application of this integrated framework, delving into its architecture, the methods employed, and the corresponding results. We share insights into the model's performance and its comparison with traditional methods. Furthermore, we provide an overview of the potential challenges and ethical considerations associated with the application of this technology in healthcare.

The journey towards an effective, efficient, and expedient stroke diagnosis and management system, marrying the power of Deep Learning and the Internet of Medical Things, promises a new dawn in healthcare. Through the lens of this research, we invite readers to envision a future where technology is not merely an adjunct but a cornerstone of patient care, enhancing

the quality of care, and improving outcomes for stroke patients worldwide. This paper aims to push the boundaries of our current understanding and application of technology in stroke management and invites the medical community to partake in this exciting journey of discovery and innovation.

II. RELATED WORKS

The fusion of Deep Learning (DL) and the Internet of Medical Things (IoMT) marks an exciting nexus of two dominant themes in recent healthcare technology research. To appreciate the novelty and value of our work, it is essential to understand the broader landscape of these areas, which this section will expound upon.

In the realm of DL, numerous studies have demonstrated its potential for image analysis in various medical fields. The convolutional neural network (CNN), a class of deep, feed-forward artificial neural networks, has been particularly instrumental in image classification tasks [9]. The advent of DL has invigorated the field of medical image analysis, pushing the boundaries of what was previously possible.

Krizhevsky et al. (2012) pioneered the application of DL in image recognition, developing a CNN model, known as AlexNet, which significantly outperformed other models in the ImageNet Large Scale Visual Recognition Challenge [10]. This seminal work laid the foundation for subsequent exploration of DL for medical imaging. For instance, Esteva et al. (2017) deployed CNNs for skin cancer diagnosis from clinical images, demonstrating a performance on par with dermatologists [11]. Further, Gulshan et al. (2016) employed a DL model to detect diabetic retinopathy and macular edema in retinal fundus photographs, meeting or exceeding the performance of human graders [12].

When it comes to stroke diagnosis and prognosis, DL has proven valuable. Havaei et al. (2017) utilized a DL-based method to segment brain tumors, highlighting the potential of DL for analyzing complex brain images [13]. Zhang et al. (2020) implemented a DL model for analyzing CT angiography and achieving accurate prediction of large-vessel occlusion strokes, underscoring the utility of DL in stroke diagnosis [14].

Yet, DL's utility in healthcare is not just confined to imaging. It has also demonstrated potential in Electronic Health Record (EHR) data analysis, predictive modeling, and health monitoring. Miotto et al. (2018) used DL to predict disease onset from EHRs, further expanding the realm of its application [15].

Parallel to DL's rise, IoMT has begun to revolutionize healthcare, promising improved patient outcomes, cost-effective care, and operational efficiency [16]. The IoMT enables interconnectivity between medical devices and healthcare IT systems, allowing for real-time patient monitoring and data collection, ultimately leading to improved clinical decision-making [17].

However, literature specifically dealing with the application of IoMT in stroke management is still sparse. The few existing studies primarily focus on IoMT's role in monitoring patients' vital parameters and rehabilitation post-stroke [18]. By

integrating the continuous monitoring of vital signs with emergency medical systems, Tang et al. (2017) demonstrated IoMT's potential to enhance pre-hospital care for stroke patients [19]. Similarly, Yan et al. (2018) developed a rehabilitation system based on IoMT, demonstrating its utility in post-stroke recovery and rehabilitation [20].

The convergence of DL and IoMT is an emerging theme in healthcare, underpinning a paradigm shift towards more integrated, data-driven patient care [21]. The fusion of these two technologies promises to unlock new levels of efficiency, precision, and patient empowerment in healthcare delivery [22]. Yet, the application of this integrated approach to stroke management, specifically the analysis of brain CT images, has remained largely unexplored, marking a gap in the literature that our research seeks to fill.

In conclusion, the amalgamation of DL and IoMT opens new horizons for stroke diagnosis and management. While both technologies have individually demonstrated their worth in healthcare, their combined application to analyze brain CT images of stroke patients is a new frontier. Our research is situated at this intersection, aiming to push the envelope further, enabling quicker, more accurate stroke diagnosis, and fostering timely intervention.

III. MATERIALS AND METHODS

The subsequent section, "Materials and Methods", forms the crux of our research, outlining the procedures, techniques, and tools employed to develop and evaluate our integrated Deep Learning and Internet of Medical Things model [23]. The fundamental aspects discussed herein include the data collection process, the architectural design of our deep learning model, the deployment of IoMT infrastructure, and the specifics of our experimental setup.

We elaborate on the dataset comprising the brain Computed Tomography (CT) images of stroke patients and the subsequent data preprocessing steps undertaken to ensure the readiness of data for model training [24]. The detailed explanation of the deep learning architecture provides a comprehensive understanding of the model's ability to decipher complex patterns in the CT images. In parallel, we delineate how the IoMT network is set up, providing an insight into the interconnected web of devices, which allows seamless and real-time exchange of crucial patient data [25].

Moreover, we shed light on the rigorous evaluation methodologies adopted to assess the performance and reliability of our proposed system. All methods are discussed in detail to ensure reproducibility of the research and to allow other researchers to leverage our work as a stepping stone for further innovations in the field.

Ultimately, the goal of this section is to provide a clear and meticulous explanation of the research methodology that led to our findings, while maintaining a scientific rigor that upholds the principles of transparency and reproducibility in academic research. This foundational knowledge will aid readers in understanding the ensuing results and discussion section, where we delve deeper into the outcomes of our research and their implications in the wider healthcare context.

Our research consists of a comprehensive stroke investigation system, ranging from detection and classification to segmentation. Early diagnosis is carried out at the IoMT level using the medical CW-4 sensor and Raspberry Pi 4 microcontroller. With the help of the medical sensor, we determine the blood flow velocity through the carotid artery in the patient at an early stage, where it either corresponds to the norm or shows deviations. In case of deviations from the norm, the patient undergoes CT/MRI diagnostics. Using CT/MRI images, the patient can perform classification through our web application, where, using a CNN model, they can obtain a result indicating the presence or absence of a stroke. If a stroke is present, the patient can further perform segmentation of the stroke lesion in the brain using a modified UNet model. Thus, the patient can receive diagnosis in several stages using our comprehensive system (Fig. 1 is Flowchart of complex system of stroke diagnosis).

A. Data

In this research, a publicly available Kaggle platform [26] was used as the dataset for classification. This dataset is divided into three groups following an 80%/20% split (training, validation, and testing) and contains 993 cases of healthy vaccinations and 610 stroke cases for the training category; 240 healthy cases and 146 stroke cases, as well as 313 healthy cases and 189 stroke cases for testing. The images in the dataset were provided as shown in Fig. 2.

We use ISLES 2018 (Ischemic Stroke Lesion Segmentation) dataset [27] for segmentation. The ISLES 2018 dataset, a key component of our research, is a robust and publicly available collection of multi-center, multi-vendor, and multi-disease stage clinical data. The dataset's diversity and size make it a compelling resource for training our deep learning model and testing its performance in real-world settings.

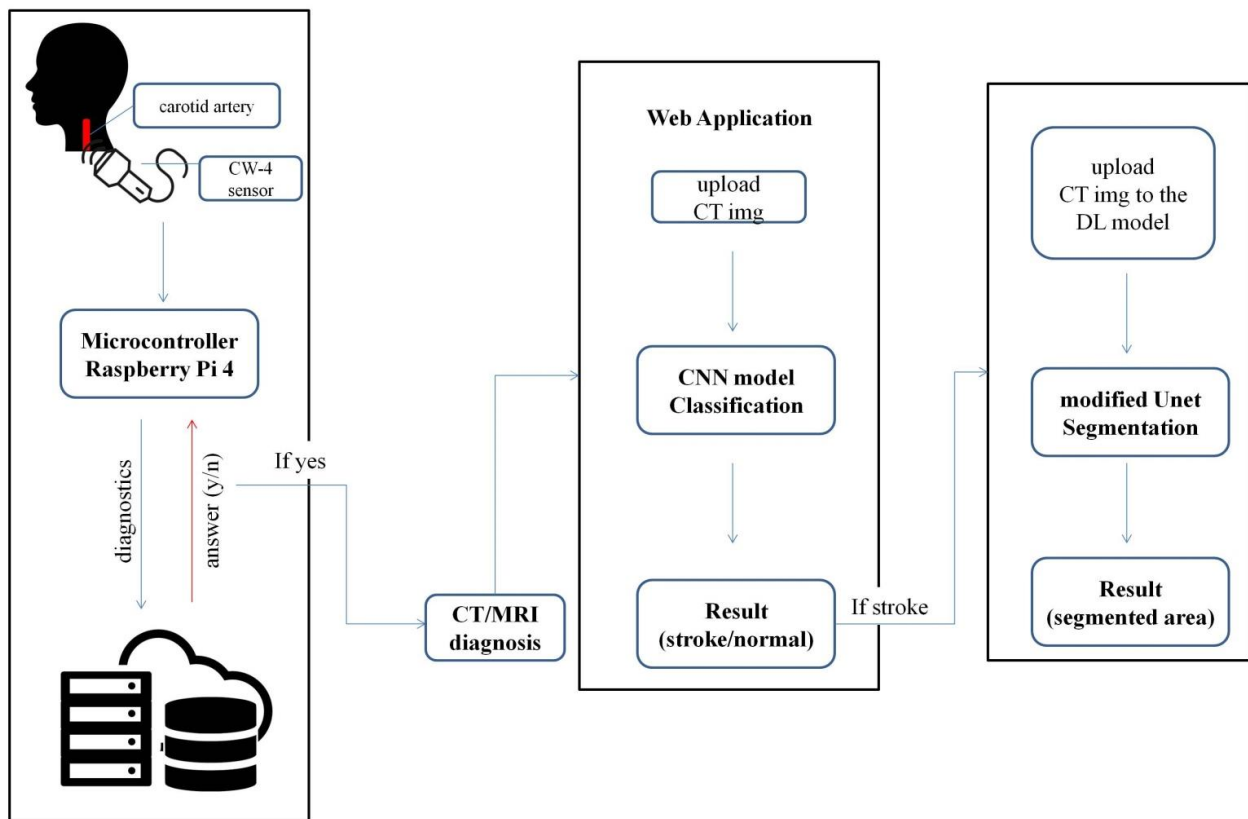


Fig. 1. Flowchart of complex system of stroke diagnosis.

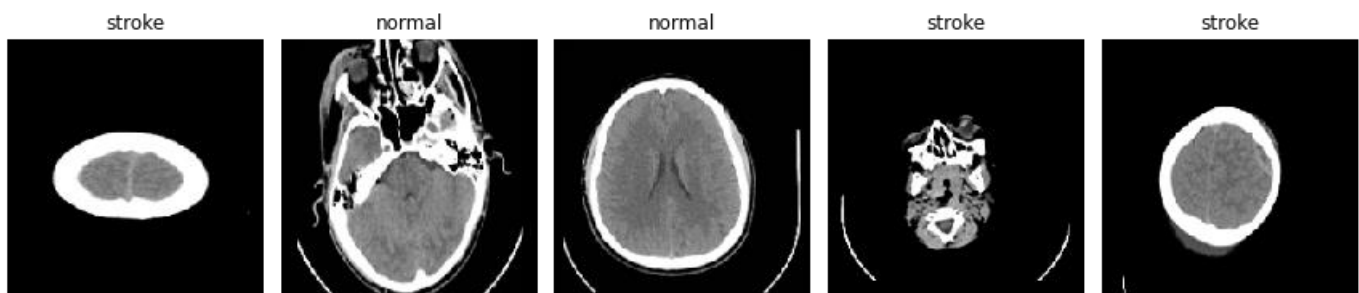


Fig. 2. Contents of the classification dataset.

ISLES 2018 contains 94 sets of computed tomography perfusion (CTP) images, with accompanying clinical metadata, sourced from multiple hospitals worldwide. Each case presents a unique story of acute ischemic stroke, offering insights into the disease's heterogeneity. The CTP scans consist of series of images, taken at different time points, which capture the progression of contrast agent through the brain vasculature, providing critical information on cerebral blood flow, blood volume, and mean transit time [28]. These image sets offer a unique opportunity to assess the impact and extent of the stroke, thus serving as a valuable ground truth for model training and validation.

The dataset is designed to ensure the balance between patient privacy and scientific value. All patient identifiers have been removed to preserve anonymity, ensuring compliance with data privacy regulations. Despite this anonymization, the dataset retains rich clinical metadata, including patients' age, sex, and stroke severity (measured by the National Institutes of Health Stroke Scale), all of which can be instrumental in informing the deep learning model's interpretations and predictions.

It is important to note that the ISLES 2018 dataset provides expert-annotated lesion segmentation masks for each CTP scan. These masks, which identify the location and extent of ischemic lesions, are a crucial component of our supervised learning approach. They allow us to train the model to recognize similar patterns in unseen CT images, and ultimately, to predict the occurrence of stroke and its impact. Fig. 3 demonstrate samples of ISLES 2018 dataset that applied in this research.

The ISLES 2018 dataset, as referenced in [29], stands as an indispensable and clinically pertinent reservoir for the iterative refinement and subsequent validation of our proposed model. It promises to underpin an enhanced degree of generalizability, making it quintessential for navigating the multifaceted and frequently intricate landscapes inherent in clinical settings. By harnessing this dataset, our endeavors transcend mere theoretical paradigms, positioning us to grapple directly with the nuanced complexities characteristic of stroke diagnostic procedures. Consequently, this deliberate engagement not only fortifies our model's robustness but also amplifies its translational potential, signifying a notable advancement in bridging the gap between academic research and its tangible clinical implementations.

B. IoMT Diagnosis

At the initial stage, early stroke diagnosis is performed using a medical ultrasonic sensor, CW-4, to determine the blood flow velocity. With the use of this sensor, the blood flow velocity through the patient's carotid artery will be measured. The obtained data will be sent over Wi-Fi to the cloud, where it will be compared with blood flow information, and the response will be sent back to the Raspberry Pi microcontroller. The first stage is mobile and portable (Fig. 1 is Flowchart of complex system of stroke diagnosis). If a deviation from the norm is detected, the patient is suggested to undergo a CT/MRI scan. Using the acquired CT/MRI results, the patient can obtain outcomes through our models for classification and segmentation without the involvement of a specialist doctor.

C. CNN Classification

Leveraging the technological capabilities of streamclip and ngrok, a Python-driven web application was meticulously crafted to facilitate CNN-based image classification. As elucidated in Fig. 1, this platform empowers users to upload cerebral images, wherein the embedded CNN model subsequently discerns between healthy and potential stroke-afflicted specimens. This model was parameterized with inputs dimensioned at (200, 200, 1), and the cumulative count of trainable parameters reached a total of 214,145. A detailed exposition of the CNN's architectural design tailored for cerebral stroke delineation is presented in Fig. 4.

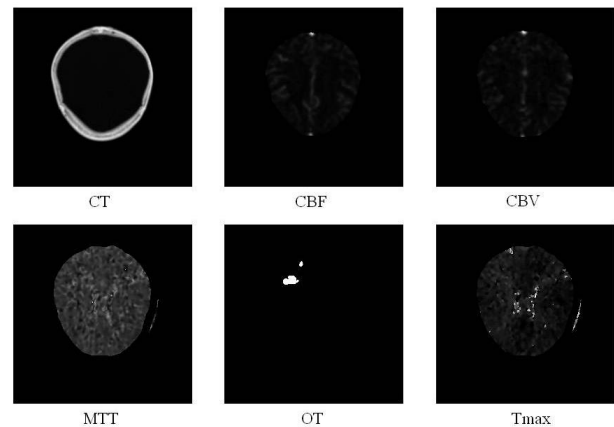


Fig. 3. Samples applied of ISLES 2018 dataset (computed tomography (CT), cerebral blood flow (CBF), cerebral blood volume (CBV), mean transit time (MTT), segmentation image (OT), tissue residue function (Tmax).

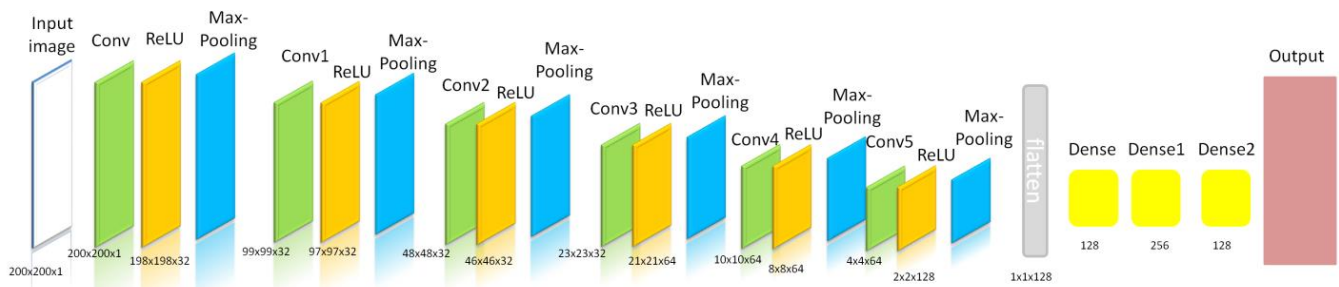


Fig. 4. Architecture of the Proposed CNN.

D. UNet Architecture

The U-Net architecture is a pivotal component of our deep learning approach, renowned for its efficacy in biomedical image segmentation. The architecture, first proposed by Ronneberger et al. in 2015, is a type of convolutional neural network (CNN) with a distinctive U-shaped design [30]. Its design accommodates a wide and varied receptive field, crucial for the detection of intricate patterns in complex images such as brain CT scans.

The U-Net model comprises two fundamental parts: the contracting (encoding) path and the expanding (decoding) path. The contracting path employs repeated applications of convolutions and max pooling to capture the context in the image, gradually reducing the spatial dimension while increasing the feature dimension. It learns low-level features at the beginning and high-level features towards the end.

The expanding path, on the other hand, performs the opposite operation. It utilizes a sequence of up-convolutions and concatenations to gradually recover the spatial dimension, making use of feature maps from the contracting path (skip-connections) to retain the precise localization information lost during contraction. This process results in a high-resolution, detailed feature map that perfectly aligns with the original image space, allowing accurate segmentation.

The U-Net architectural paradigm, distinguished by its proficiency in seamlessly integrating localized and expansive image insights, stands out as particularly germane for the intricate analysis of brain CT scans. Given the inherent intricacies of these images, a meticulous observation of nuanced elements becomes imperative to accurately detect the often understated manifestations of a stroke. This model astutely maintains equilibrium between assimilating broader contextual nuances and precisely demarcating the spatial positioning of salient features. Such an adept balance underscores the U-Net model's pivotal and irreplaceable contribution to our scholarly investigation.

E. Proposed Model

In our study, we adapted the UNet architecture to enhance its accuracy in segmenting stroke cases in computed tomography. We introduced modifications to the classic 3D UNet model using various techniques, including data augmentation, dropout, the Adam optimization algorithm, l2 regularization, and instance normalization. Each of these methodologies brings its own undeniable benefits to the table.

Fig. 6 illustrates the proposed UNet architecture. In this neural network structure, every neuron is linked to all preceding layer neurons, each linkage carrying its unique weight factor. Within a convolutional neural network, a small weight matrix—utilized in convolution operations—is slid across the entire processed layer (at the network's input, directly along with the input image). The convolution layer aggregates the results of the element-wise multiplication of each image segment with the convolution kernel matrix. The weight coefficients of the convolution kernel remain undetermined and are set during the learning process [31].

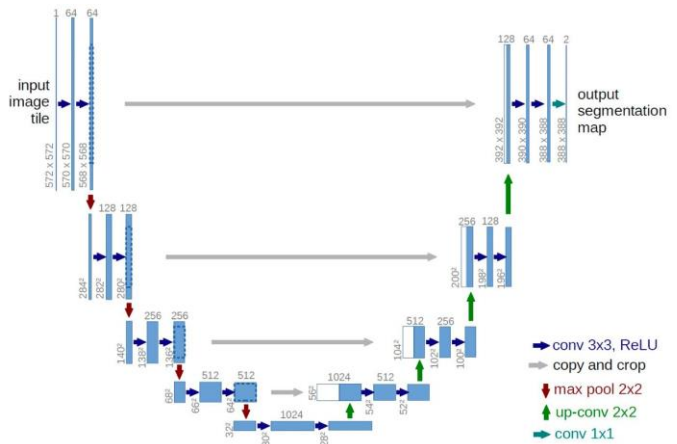


Fig. 5. UNet architecture.

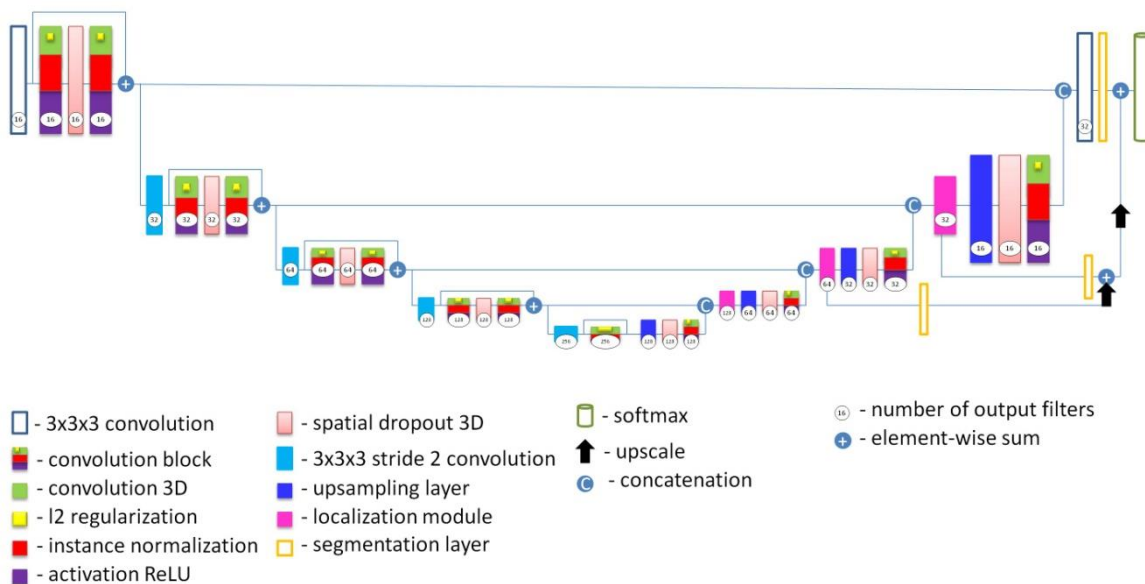


Fig. 6. Proposed enhanced UNet architecture for stroke segmentation.

Model data sizes were as follows: input shape = [5, 128, 128, 32], weight decay=0. For comparison, the classic 3D UNet was evaluated over 200 epochs, while the proposed 3D UNet model ran for 650 epochs. Upon evaluation, the classic 3D UNet model achieved a dice/f1 score of 48%, precision of 39%, recall/sensitivity of 99%, and a Jaccard index of 35% during training, while the proposed model received scores of 90%, 83%, 93%, and 89% on the same metrics, respectively. In terms of testing, the classic 3D UNet model yielded a dice/f1 score of 36%, precision of 38%, recall/sensitivity of 37%, and a Jaccard index of 32%, while the proposed model achieved a dice/f1 score of 58%, precision of 68%, recall/sensitivity of 60%, and a Jaccard index of 66%.

IV. EVALUATION METRICS

In the process of evaluating the efficacy of the proposed model, we leverage several evaluation metrics.

Accuracy is an indicator that illustrates the accuracy rate of the model prediction across all parameters. It is measured as the percentage of correct predictions made by the model. This is particularly helpful in situations in which all of the classes are of similar importance. The formula for determining it is the ratio of the number of accurate forecasts to the total number of predictions made. In fact, this is the probability that the class will be predicted correctly. Eq. (1) demonstrates formula of accuracy.

$$Accuracy(a) = \frac{\sum_{i=1}^N \mathbb{1}[a(x_i) = y_i]}{N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Here, TP is true positives, TN is true negatives, FP is false positives, FN is false negatives.

The Dice Similarity Coefficient (DSC), also known as the Sørensen–Dice index, is a statistical measure used extensively in image segmentation tasks, particularly for evaluating the performance of image classification models [32]. It quantifies the overlap between two binary images, usually the ground truth and the predicted output. Computationally, DSC is the twice the area of overlap between the two images divided by the total number of pixels in both images. A DSC score of 1 represents perfect agreement, while a score of 0 denotes no overlap. In medical image analysis, it aids in assessing the quality of segmentation models.

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

The Jaccard index, or Jaccard similarity coefficient, is a statistical measure widely used for comparing the similarity and diversity of sample sets. In the context of image segmentation and classification, it assesses the overlap between the predicted output and the ground truth. Computationally, it is the intersection (area of overlap) divided by the union (total area) of two binary images. The Jaccard index ranges from 0 to 1, where a score of 1 indicates perfect overlap and a score of 0 suggests no overlap. It is a critical evaluation metric in various

domains, including medical image analysis, where it quantifies the accuracy of segmentation models.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

Precision, often referred to as the positive predictive value, is a key performance metric in statistical classification tasks, measuring the proportion of correctly identified positive instances out of all instances that the model predicted as positive. It evaluates the exactness or quality of a classifier by highlighting its false positive rate [33]. A high precision score indicates fewer false positives, meaning the model has accurately predicted the positive instances. However, precision alone doesn't account for false negatives (actual positives predicted as negative), and thus, it's usually used alongside other metrics like recall and F1-score to provide a comprehensive model evaluation.

$$precision = \frac{TP}{TP + FP} \quad (4)$$

Recall, also known as sensitivity or true positive rate, is a critical performance metric in statistical classification models [34]. It measures the proportion of actual positive cases that are correctly identified by the model. In essence, recall gauges a model's ability to find all the relevant instances within a dataset. A high recall indicates a low rate of false negatives, meaning the model has effectively captured the positive instances. However, it's worth noting that recall doesn't account for false positives (predicted positives that are actually negatives). As such, it's commonly used alongside precision and F1-score for a more holistic evaluation of a model's performance.

$$recall = \frac{TP}{TP + FN} \quad (5)$$

The harmonic mean between accuracy and completeness is denoted by the letter F-measure. If either accuracy or completeness trend towards 0, then so does this metric. Eq. (6) demonstrate formula of the F-measure evaluation parameter.

$$Fmeasure = \frac{2 Precision \bullet Recall}{Precision + Recall} \quad (6)$$

V. EXPERIMENTAL RESULTS

To demonstrate the functionality of the CNN classification model, a web application was created using Python along with ngrok and streamline. As shown in Fig. 7, a brain image is uploaded to the web application. Subsequently, the CNN model performs classification and provides a response based on the model's results. In this instance, the model correctly classified the image as normal with an accuracy of 79%.

Our experimental work leveraged the established U-Net architecture and the ISLES 2018 dataset. We carried out the practical portion of the experiment using the Tensorflow library within the Google Colab environment.

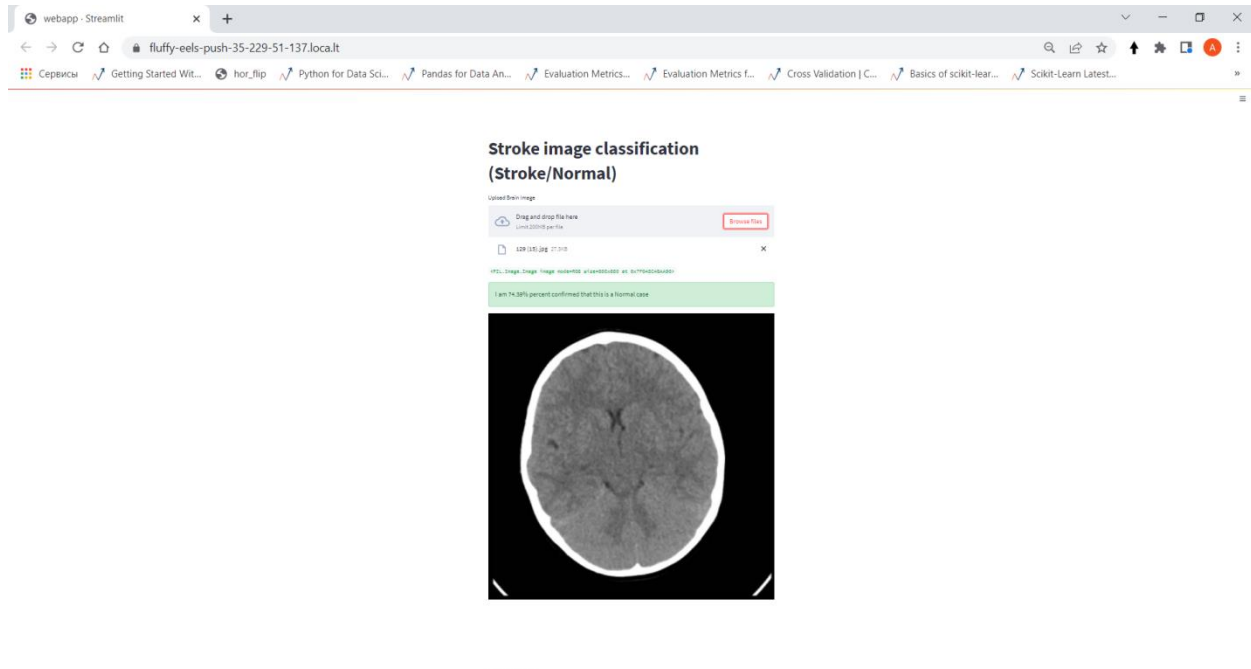


Fig. 7. Stroke classification web app on CNN model.

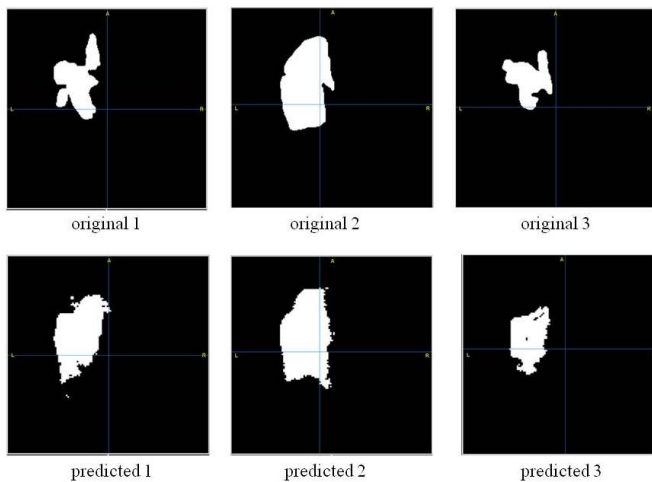


Fig. 8. Brain stroke segmentation results. (First line images are original images; Second line images are the images after applying segmentation).

Fig. 8 depict three instances of original segmented images alongside their predicted counterparts according to the proposed UNet model. Initially, the original images exhibited superior quality. However, to reduce the number of neural network parameters for computations in Google Colab, we had to resort to image compression [35]. Consequently, due to the degradation in image quality, the predicted images show deviations compared to their original versions. Nevertheless, the segmentation of the stroke lesion area was correctly identified and delineated.

VI. DISCUSSION

In this study, we have presented a comprehensive investigation into a deep learning approach using the Internet of Medical Things (IoMT) for analyzing computed

tomography (CT) brain images of stroke patients [36]. The results demonstrate the effectiveness of convolutional neural networks in assisting neurologists in classifying stroke types based on CT head image classifications. Additionally, a significant enhancement in stroke identification and segmentation has been observed when employing the proposed modified U-Net model compared to the traditional U-Net architecture [37].

The significance of this research lies not only in the achieved outcomes but also in the methodologies and techniques employed. The utilization of the 2018 ISLES dataset, one of the most comprehensive stroke visualization datasets, provides a robust foundation for analysis [38]. Moreover, the use of an enhanced deep learning model to tackle such a complex and critical task underscores the potential that artificial intelligence and IoMT hold in the realm of healthcare.

Our proposed model modifies the U-Net architecture to achieve higher segmentation accuracy of stroke regions on CT images. Yet, this study also encountered certain limitations. For example, the compression of high-quality images for Google Colab's computation requirements resulted in degraded image quality, impacting the accuracy of the predicted images. This underscores the need for advanced computational capabilities to handle high-resolution medical imaging data without compromising on quality, thus retaining the critical details needed for accurate diagnosis.

However, even with these challenges, the proposed model achieved promising results. Compared to the traditional 3D UNet model, our modified model demonstrated higher Dice/F1 scores, precision, recall/sensitivity, and Jaccard index, both during training and in test results. These metrics provide clear evidence of the model's superior performance in segmenting stroke regions in CT images.

As with any study, the future direction of this research hinges on the lessons learned. The noted constraints will serve as considerations in future research, especially in relation to data and computation requirements. The successes of this study also pave the way for more advanced deep learning and IoMT applications in medical imaging analysis [39]. Specifically, more complex and adaptable deep learning models can be explored for more precise and reliable results.

To sum up, this study illustrates the power of AI in the IoMT context to accurately analyze CT images of stroke patients. Despite its limitations, the complex system of stroke diagnosing show considerable promise and lays the groundwork for future studies in this field. It emphasizes the need for continued research and development in this area to fully realize the potential of deep learning and IoMT in transforming stroke diagnosis and treatment. As the field continues to evolve, these technological advancements will undoubtedly play a pivotal role in enhancing patient care, improving outcomes, and ultimately, saving lives.

VII. CONCLUSION

In conclusion, this research provides valuable insights into the integration of deep learning and the Internet of Medical Things (IoMT) within medical imaging analysis, specifically focusing on Brain Computed Tomography (CT) images of stroke patients. Leveraging the advantages of IoMT, CNN classification, and the modified U-Net model, substantial progress has been made in enhancing the accuracy of stroke detection, classification, and segmentation. This, in turn, contributes to informing the decision-making processes in stroke treatment.

However, the research also identified computational constraints and image quality degradation as challenges that need addressing in future studies. The computational power required for high-resolution medical image processing and deep learning model training, along with the necessity for maintaining the quality of original images, are aspects that future work must address to harness the full potential of AI and IoMT in healthcare.

Despite these challenges, the research underscores the transformative potential of deep learning and IoMT in healthcare. It highlights the ability of advanced AI models to deliver accurate and precise results that can potentially revolutionize the process of diagnosing and treating strokes. This study, therefore, sets the stage for further exploration and enhancement of AI models in medical imaging analysis.

In closing, the application of deep learning and IoMT in the field of medical imaging is a burgeoning area of study. With continued research and technological advancements, we are optimistic about the prospects of these tools in bringing about a paradigm shift in the diagnosis and treatment of critical health conditions like strokes, ultimately improving patient outcomes.

REFERENCES

- [1] M. M. Yapici, R. Karakis and K. Gurkahraman, "Improving brain tumor classification with deep learning using synthetic data," *Computers, Materials & Continua*, vol. 74, no.3, pp. 5049–5067, 2023.
- [2] Latif, A. I., Daher, A. M., Suliman, A., Mahdi, O. A., & Othman, M. (2019). Feasibility of Internet of Things application for real-time healthcare for Malaysian pilgrims. *Journal of Computational and Theoretical Nanoscience*, 16(3), 1169-1181.
- [3] Xu, Y., Souza, L. F., Silva, I. C., Marques, A. G., Silva, F. H., Nunes, V. X., ... & Rebouças Filho, P. P. (2021). A soft computing automatic based in deep learning with use of fine-tuning for pulmonary segmentation in computed tomography images. *Applied Soft Computing*, 112, 107810.
- [4] Chen, Y. T., Chen, Y. L., Chen, Y. Y., Huang, Y. T., Wong, H. F., Yan, J. L., & Wang, J. J. (2022). Deep learning-based brain computed tomography image classification with hyperparameter optimization through transfer learning for stroke. *Diagnostics*, 12(4), 807.
- [5] Salleh, N. S. M., Suliman, A., & Jørgensen, B. N. (2020, August). A systematic literature review of machine learning methods for short-term electricity forecasting. In *2020 8th International conference on information technology and multimedia (ICIMU)* (pp. 409-414). IEEE.
- [6] Chandrabhatla, A. S., Kuo, E. A., Sokolowski, J. D., Kellogg, R. T., Park, M., & Mastorakos, P. (2023). Artificial Intelligence and Machine Learning in the Diagnosis and Management of Stroke: A Narrative Review of United States Food and Drug Administration-Approved Technologies. *Journal of Clinical Medicine*, 12(11), 3755.
- [7] Omarov, B., Altayeva, A., Turganbayeva, A., Abdulkarimova, G., Gusmanova, F., Sarbasova, A., ... & Omarov, N. (2019). Agent based modeling of smart grids in smart cities. In *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5* (pp. 3-13). Springer International Publishing.
- [8] Son, H., Lee, S., Kim, K., Koo, K. I., & Hwang, C. H. (2022). Deep learning-based quantitative estimation of lymphedema-induced fibrosis using three-dimensional computed tomography images. *Scientific Reports*, 12(1), 15371.
- [9] Yeo, M., Tahayori, B., Kok, H. K., Maingard, J., Kutaiba, N., Russell, J., ... & Asadi, H. (2021). Review of deep learning algorithms for the automatic detection of intracranial hemorrhages on computed tomography head imaging. *Journal of neurointerventional surgery*, 13(4), 369-378.
- [10] Salleh, N. S. M., Suliman, A., & Ahmad, A. R. (2011, November). Parallel execution of distributed SVM using MPI (CoDLib). In *ICIMU 2011: Proceedings of the 5th international Conference on Information Technology & Multimedia* (pp. 1-4). IEEE.
- [11] Al Rub, S. A., Alaiad, A., Hmeidi, I., Quwaider, M., & Alzoubi, O. (2023). Hydrocephalus classification in brain computed tomography medical images using deep learning. *Simulation Modelling Practice and Theory*, 123, 102705.
- [12] Chavva, I. R., Crawford, A. L., Mazurek, M. H., Yuen, M. M., Prabhat, A. M., Payabvash, S., ... & Sheth, K. N. (2022). Deep learning applications for acute stroke management. *Annals of Neurology*, 92(4), 574-587.
- [13] Neethi, A. S., Niyas, S., Kannath, S. K., Mathew, J., Anzar, A. M., & Rajan, J. (2022). Stroke classification from computed tomography scans using 3d convolutional neural network. *Biomedical Signal Processing and Control*, 76, 103720.
- [14] Al-Mekhlafi, Z. G., Senan, E. M., Rassem, T. H., Mohammed, B. A., Makbol, N. M., Alanazi, A. A., ... & Ghaleb, F. A. (2022). Deep learning and machine learning for early detection of stroke and haemorrhage. *Computers, Materials and Continua*, 72(1), 775-796.
- [15] Manickam, P., Mariappan, S. A., Murugesan, S. M., Hansda, S., Kaushik, A., Shinde, R., & Thipperudraswamy, S. P. (2022). Artificial intelligence (AI) and internet of medical things (IoMT) assisted biomedical systems for intelligent healthcare. *Biosensors*, 12(8), 562.
- [16] A. Altayeva, B. Omarov, H.C. Jeong and Y.I. Cho, "Multi-step face recognition for improving face detection and recognition rate", *Far East Journal of Electronics and Communications*, vol. 16, no. 3, pp. 471-491, 2016.
- [17] Woźniak, M., Siłka, J., & Wiczorek, M. (2021). Deep neural network correlation learning mechanism for CT brain tumor detection. *Neural Computing and Applications*, 1-16.
- [18] Qiu, W., Kuang, H., Ospel, J. M., Hill, M. D., Demchuk, A. M., Goyal, M., & Menon, B. K. (2021). Automated prediction of ischemic brain tissue fate from multiphase computed tomographic angiography in

- patients with acute ischemic stroke using machine learning. *Journal of stroke*, 23(2), 234-243.
- [19] Lin, S. Y., Chiang, P. L., Chen, M. H., Lee, M. Y., Lin, W. C., & Chen, Y. S. (2023). DGA3-Net: A parameter-efficient deep learning model for ASPECTS assessment for acute ischemic stroke using non-contrast computed tomography. *NeuroImage: Clinical*, 38, 103441.
- [20] Mouli, D. V. C. (2023, March). Artificial Intelligence Enabled Framework for Automatic Brain Stroke Detection. In 2023 10th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 659-664). IEEE.
- [21] Foroushani, H. M., Hamzehloo, A., Kumar, A., Chen, Y., Heitsch, L., Slowik, A., ... & Dhar, R. (2022). Accelerating prediction of malignant cerebral edema after ischemic stroke with automated image analysis and explainable neural networks. *Neurocritical Care*, 36(2), 471-482.
- [22] Saxena, S., Jena, B., Mohapatra, B., Gupta, N., Kalra, M., Scartozzi, M., ... & Suri, J. S. (2023). Fused deep learning paradigm for the prediction of o6-methylguanine-DNA methyltransferase genotype in glioblastoma patients: A neuro-oncological investigation. *Computers in Biology and Medicine*, 153, 106492.
- [23] Chen, I. E., Tsui, B., Zhang, H., Qiao, J. X., Hsu, W., Nour, M., ... & Nael, K. (2022). Automated estimation of ischemic core volume on noncontrast-enhanced CT via machine learning. *Interventional Neuroradiology*, 15910199221145487.
- [24] Khan, M., Shah, P. M., Khan, I. A., Islam, S. U., Ahmad, Z., Khan, F., & Lee, Y. (2023). IoMT-Enabled Computer-Aided Diagnosis of Pulmonary Embolism from Computed Tomography Scans Using Deep Learning. *Sensors*, 23(3), 1471.
- [25] Wang, L. J., Zhai, P. Q., Xue, L. L., Shi, C. Y., Zhang, Q., & Zhang, H. (2023). Machine learning-based identification of symptomatic carotid atherosclerotic plaques with dual-energy computed tomography angiography. *Journal of Stroke and Cerebrovascular Diseases*, 32(8), 107209.
- [26] Brain Stroke CT Image Dataset, 2022. [Online]. Available: <https://www.kaggle.com/datasets/afriDirahman/brain-stroke-ct-image-dataset>.
- [27] Subudhi, A., Dash, P., Mohapatra, M., Tan, R. S., Acharya, U. R., & Sabut, S. (2022). Application of Machine Learning Techniques for Characterization of Ischemic Stroke with MRI Images: A Review. *Diagnostics*, 12(10), 2535.
- [28] Kawai, Y., Kogeichi, Y., Yamamoto, K., Miyazaki, K., Asai, H., & Fukushima, H. (2023). Explainable artificial intelligence-based prediction of poor neurological outcome from head computed tomography in the immediate post-resuscitation phase. *Scientific Reports*, 13(1), 5759.
- [29] Elbagoury, B. M., Vladareanu, L., Vlădăreanu, V., Salem, A. B., Travediu, A. M., & Roushdy, M. I. (2023). A Hybrid Stacked CNN and Residual Feedback GMDH-LSTM Deep Learning Model for Stroke Prediction Applied on Mobile AI Smart Hospital Platform. *Sensors*, 23(7), 3500.
- [30] Patel, T. R., Patel, A., Veeturi, S. S., Shah, M., Waqas, M., Monteiro, A., ... & Tutino, V. M. (2023). Evaluating a 3D deep learning pipeline for cerebral vessel and intracranial aneurysm segmentation from computed tomography angiography–digital subtraction angiography image pairs. *Neurosurgical Focus*, 54(6), E13.
- [31] Hsu, K., Yuh, D. Y., Lin, S. C., Lyu, P. S., Pan, G. X., Zhuang, Y. C., ... & Juan, C. J. (2022). Improving performance of deep learning models using 3.5 D U-Net via majority voting for tooth segmentation on cone beam computed tomography. *Scientific Reports*, 12(1), 19809.
- [32] Zhou, Q., Zhao, R., Hu, Y., Wang, J., & Zhou, R. (2023). Hierarchical Hybrid Networks for Automatic Pulmonary Blood Vessel Segmentation in Computed Tomography Images. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.
- [33] Gudigar, A., Raghavendra, U., Hegde, A., Menon, G. R., Molinari, F., Ciaccio, E. J., & Acharya, U. R. (2021). Automated detection and screening of traumatic brain injury (TBI) using computed tomography images: a comprehensive review and future perspectives. *International journal of environmental research and public health*, 18(12), 6499.
- [34] Rana, A., Dumka, A., Singh, R., Panda, M. K., & Priyadarshi, N. (2022). A Computerized Analysis with Machine Learning Techniques for the Diagnosis of Parkinson's Disease: Past Studies and Future Perspectives. *Diagnostics*, 12(11), 2708.
- [35] Omarov, B., Altayeva, A., Suleimenov, Z., Im Cho, Y., & Omarov, B. (2017, April). Design of fuzzy logic based controller for energy efficient operation in smart buildings. In 2017 First IEEE International Conference on Robotic Computing (IRC) (pp. 346-351). IEEE.
- [36] Raghavendra, U., Gudigar, A., Paul, A., Goutham, T. S., Inamdar, M. A., Hegde, A., ... & Acharya, U. R. (2023). Brain tumor detection and screening using artificial intelligence techniques: Current trends and future perspectives. *Computers in Biology and Medicine*, 107063.
- [37] Heo, S., Ha, J., Jung, W., Yoo, S., Song, Y., Kim, T., & Cha, W. C. (2022). Decision effect of a deep-learning model to assist a head computed tomography order for pediatric traumatic brain injury. *Scientific Reports*, 12(1), 12454.
- [38] Li, Y. L., Chen, C., Zhang, L. J., Zheng, Y. N., Lv, X. N., Zhao, L. B., ... & Lv, F. J. (2023). Prediction of Early Perihematomal Edema Expansion Based on Noncontrast Computed Tomography Radiomics and Machine Learning in Intracerebral Hemorrhage. *World Neurosurgery*.
- [39] Salehinejad, H., Kitamura, J., Ditkofsky, N., Lin, A., Bharatha, A., Suthiphosuwat, S., ... & Colak, E. (2021). A real-world demonstration of machine learning generalizability in the detection of intracranial hemorrhage on head computerized tomography. *Scientific Reports*, 11(1), 17051.

Chatbot Program for Proposed Requirements in Korean Problem Specification Document

Young Yun Baek¹, Soojin Park², Young B. Park³

Dept. of Software Science, Dankook University, Seoul, South Korea^{1,3}

Graduate School of Management of Technology, Sogang University, Seoul, South Korea²

Abstract—In software engineering, requirement analysis is a crucial task throughout the entire process and holds significant importance. However, factors contributing to the failure of requirement analysis include communication breakdowns, divergent interpretations of requirements, and inadequate execution of requirements. To address these issues, a proposed approach involves utilizing NLP machine learning within Korean requirement documents to generate knowledge-based data and deduce actors and actions using natural language processing knowledge-based information. Actors and actions derived are then structured into a hierarchy of sentences through clustering, establishing a conceptual hierarchy between sentences. This is transformed into ontology data, resulting in the ultimate requirement list. A chatbot system provides users with the derived system event list, generating requirement diagrams and specification documents. Users can refer to the chatbot system's outputs to extract requirements. In this paper, the feasibility of this approach is demonstrated by applying it to a case involving Korean-language requirements for course enrollment.

Keywords—Requirement engineering; NLP machine learning; clustering; Korean document; chatbot

I. INTRODUCTION

In software engineering, requirements analysis is the process of eliciting, analyzing, specifying, and validating the requirements that must be satisfied in software development. It is a crucial activity in the early stages of development and has a significant impact on the design and other phases of the software development lifecycle. Insufficient requirements analysis can lead to project failures, while a proper requirements elicitation process serves as the foundation for overall software product quality. Therefore, it is necessary to perform requirements analysis in a thorough and specific manner, considering its importance and high impact on the overall success or failure of a project [1].

Furthermore, requirements analysis represents the client's needs, contractual obligations, standard specifications, and documented information in software development. It expresses the conditions, functionalities, or capabilities that the development program must perform. It helps identify the desired features, goals, constraints, and other necessary information from the users for system development. Thus, requirements analysis is crucial in creating software that meets user expectations and is considered the most critical process in the software development life cycle [2],[3].

However, extracting clear requirements from problem statements and capturing requirements that encompass the

overall software can still be challenging and complex due to the diverse values, attitudes, behavioral norms, beliefs, and communication of stakeholders, who have different perspectives [4]. Another reason is that problem statements are written in natural language. Natural language descriptions can be interpreted differently by individuals, and the same word can have different meanings. Furthermore, as the field of software expands and becomes more complex, analyzing clear requirements becomes difficult, and it requires fundamental knowledge to understand the system. Additionally, since humans are involved in the process of extracting these requirements, there are limitations in consistency and analysis due to the subjective nature of human work.

In this paper, to address these challenges, natural language processing and NLP artificial intelligence analysis are employed to extract key words from problem specification documents in the document phase. By extracting actor-behavior relationships and hierarchical data through clustering, sequential analysis of sentences and hierarchical analysis data are obtained, which are then structured into ontology to be used as foundational data for a chatbot. This approach allows for the extraction of consistent requirements from complex problem specification documents and proposes a method to resolve human effort by recommending and providing a list of requirements to users through question-and-answer interactions with the chatbot.

II. RELATED WORK

A. Sequential Analysis for Requirement Classification and Word Entity Extraction

In Wang et al. [5], they addressed the problem of modeling the interaction between the physical environment and users in model generation for testing and validation in NLR (Natural Language Requirements) by using NLP techniques and model mapping rules to identify model elements. Jahan et al. [4] recognized the importance of automation in behavior modeling and proposed an automated approach for constructing SD (System Design) from natural language-written use case scenarios to bridge some of the gaps in the literature. Limaylla-Lunarejo et al. [6] applied machine learning algorithms combined with natural language processing to classify software requirements into functional and non-functional categories. Koscinski et al. [7] automated the formalization of NL (Natural Language) requirements by using IE (Information Extraction) techniques to extract structured information from NL SysRS (Natural Language System Requirements) data. Güneş et al. [8] automatically generated and visualized goal models based on

user stories using a natural language processing pipeline and heuristics. Tobias et al. [9] proposed an approach called NoBERT, which utilizes the fine-tuning mechanism of BERT for classifying requirements. Saini et al. [10] proposed an automated approach that combines NLP and ML techniques to extract domain models. Tiwari et al. [11] proposed an approach using entity recognition NLP techniques to identify use case names and actor names in text-based requirements specifications. Imam et al. [12] proposed a method using SVM (Support Vector Machine) for recognizing use case entities in unstructured sentences.

These methods have strengths in sequential analysis for requirement classification and actor behavior extraction. However, they have limitations in establishing the relationships between related requirements. While it is possible to classify and extract requirements within a single sentence during requirement analysis, it is challenging to understand the overall flow and structure of the entire set of requirements, which is a drawback in grasping the flow of requirements.

B. Dependency Analysis of Requirements through Hierarchical Analysis

In Deshpande et al. [13], requirement dependencies were extracted using domain ontology and labeling learning. Zhang et al. [14] proposed an automated requirement term extraction and ranking framework based on a graph-based ranking algorithm. Wardhana et al. [15] suggested using ontology to represent the semantic context in system design.

These hierarchical analysis methods have the advantage of capturing the dependencies and hierarchical relationships between requirements by identifying the parent-child relationships among them. However, when performing hierarchical analysis alone, it is limited to simple comparisons between requirements, and it cannot analyze the relationships between words within sentences, which is a drawback.

III. RE CHATBOT SYSTEM CONFIGURATION

A. System Configuration

In [1] framework, clustering was added to enable sequential analysis and hierarchical analysis of sentences, which were then applied to ontology construction. By utilizing hierarchical analysis in ontology construction, it becomes possible to identify higher-level concepts of requirements between sentences. This allows for both word-centric sequential analysis and sentence-centric hierarchical analysis. By understanding the relationships between sentences, beyond simple sequential analysis for classification, it is possible to provide requirements that are prioritized and classified based on their inter-sentence dependencies.

Each sentence in the problem specification document is subjected to morphological analysis and BERT Q&A operations to generate the underlying knowledge data in advance. The system possesses data for sequential analysis through the aforementioned process, which is then combined with the results of clustered sentence data to construct ontology. Through the ontology, entities such as actors,

actions, events, and systems are extracted. The extracted relationship data and ontology data exist as base data in the chatbot program. Users go through the process of final requirement specification by interacting with the chatbot through requirement Q&A sessions. Users can review the lists provided by the chatbot Q&A and use them as a reference to compose their requirements. The new system structure is depicted in Fig. 1.

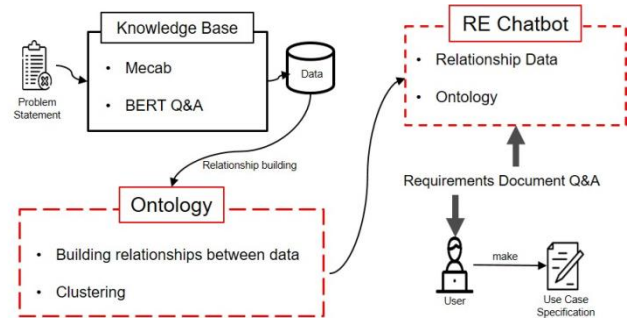


Fig. 1. System overview of process of creating components for writing the use case specification in the problem description document

B. Mecab Morphological Analysis, BERT Q&A Knowledge Base

The process of constructing the knowledge base data is described in [1] following the approach. The problem statement is preprocessed using Mecab and BERT Q&A to perform morphological analysis and sequential analysis of the sentences, respectively. The Mecab morphological analysis method for constructing the knowledge base data from the problem statement is shown in Fig. 2, while the BERT Q&A analysis method is illustrated in Fig. 3.

C. Clustering Configuration

For clustering, two types of clustering were performed to cluster sentences. First, a tf-idf analysis was performed using only the subject words within the sentences. By excluding unnecessary words such as particles, adverbs, adjectives, etc., it becomes possible to analyze the core actors and actions in Korean sentences. Furthermore, by analyzing tf-idf based on the subject words, which focuses on the subject words, clustering can be performed. Since it is subject-word-centered clustering, it is possible to perform sequential analysis with the results of morphological analysis and sequential analysis of the sentences, allowing for clustered analysis.

Next, the results of tf-idf analysis are used to perform DBSCAN clustering and Agglomerative clustering. DBSCAN clustering is applied to obtain clustered results of the problem statement text based on subject words. This allows for the analysis of similar sentences that include a specific subject word and the analysis of similar subject words. Agglomerative clustering is performed to obtain hierarchical clustering results among sentences. This enables the analysis of the hierarchical relationships of a specific sentence with other sentences. Through these two clustering methods, the sentences are analyzed in a hierarchical structure, allowing for the extraction of sentences with hierarchical relationships.

- **Question:** Morphological analysis extracts words whose morpheme classification has the order "XSV => EC", "VV+ETM => NNB", and "XSV+ETM => NNB"
- **Extraction Subject:** When combining subject words + subject postposition, a match is produced by comparing them with space-writing words. The reason for extracting the subject as described above is performed to reduce the operating time that occurs when writing questions of BERT Q&A.
- **Extraction Object:** When combining subject word + object postposition, a match is generated by comparing it with a space-writing word. The reason for extracting the subject as described above is performed to reduce the operating time that occurs when writing questions of BERT Q&A.
- **Extraction Question Behavior:** Morphology classification in question words is classified according to "EC" and "NNB" to generate question history and compare it with writing words to produce matching words. The reason for extracting the subject as described above is performed to reduce the operating time that occurs when writing questions of BERT Q&A.
- **Average Word Length:** Extract to limit the length of response sentences that come out through BERT Q&A, and make the average word length based on spacing words.

Fig. 2. Morphological analysis algorithm of the RE chatbot system.

D. Ontology Configuration

Based on the extracted subject words from the pre-constructed knowledge base data, the subject part of the ontology is formed. This subject part consists of words that are used as actors in the requirements. Subsequently, the predicate part and the object part are written with the subject words as the focus. Next, the necessary steps are taken to construct the predicate part and the object part. The following is the method for constructing the predicate part and the object part, which are essential for the operation of the entire system.

To express the actions of actor words, the predicate part of the ontology is constructed as "Action". Next, to construct the object part of this predicate, the actions that actors can perform within the system are extracted from the knowledge base data and used as objects. This object part consists of sentences that are used as actions in the requirements analysis.

To construct associated words for actor words, the predicate part of the ontology is structured as "Associated Words." Then, to construct the object part of this predicate, words associated with the subject part are extracted from the clustering results and used as objects. This object part is used in requirements analysis to analyze similar words centered around a given word.

To create associated sentences for the word "actor," the ontology's descriptors are structured as "associated sentences." Next, to form the purpose of these descriptors, sentences containing subject words are extracted from the clustering results and used as the purpose. This purpose is employed to provide similar requirement sentences centered around the main sentence when inputting key sentences during requirement analysis.

- **Questions:** Extract Subject + Question Death + Object postposition + Extract Question Behavior
- The generated question is selected as the question content of the BERT model, and questions are asked based on the entire sentence. As a result, BERT Q&A uses the results found in the response. The following preprocessing is performed for the results answered.
- Remove [SEP] if it contains [SEP] for the response sentence.
 - If the response sentence contains [CLS] and [UNK], then the response sentence is not used.
 - If the response sentence exceeds the average word length, then do not use the response sentence.

Fig. 3. BERT Q&A analysis algorithm of the RE chatbot system.

E. RE Chatbot Configuration

The chatbot possesses the constructed ontology file from the previous steps, along with the clustered sentence content and hierarchy information as its data. The chatbot reads the constructed ontology file and clustering information to generate actor and action information based on the user's input of key sentences and words. It provides a prioritized list of requirements as a result. Users can utilize the requirement list provided by the chatbot for their analysis. The chatbot algorithm used by the user is illustrated in Fig. 4.

- Start
1. RE Chatbot asks the user for a requirement topic and related words.
 2. User enters requirement topic and related words in RE Chatbot.
 3. RE Chatbot searches the ontology description for the requirements topic entered by the user, checks it, and searches the relevant subject book.
 4. RE Chatbot selects the main word by comparing the subject word searched in step 3 with the related word, and based on this, searches in the clustering results.
 5. RE Chatbot selects a sentence that contains the remaining related words from the clustering results retrieved in step 4, excluding the main words.
 6. RE Chatbot constructs the requirement sentence according to the priority classified by the clustering result from the sentence selected in step 5.
 7. RE Chatbot shows the result of step 6 to the user and repeats step 1 again.
- End

Fig. 4. Chatbot operation algorithm of the RE Chatbot system.

IV. CASE STUDY APPLYING REQUIREMENTS DOCUMENT

To validate the proposed system, a selection of requirements documents, specifically those related to course enrollment, was chosen. These requirements were used to conduct the validation of the system by applying the suggested requirement diagram and requirement specification automatic generation system outlined in the paper.

A. Clustering

Clustering was performed using two approaches to cluster the sentences. First, the sentences were analyzed using tf-idf based on the extracted subject words within each sentence. DBSCAN clustering was then applied to this dataset, resulting in sentence clustering based on subject words. Similarly, the subject words within the sentences were extracted and tf-idf analysis was performed. Agglomerative clustering was applied to this dataset to achieve sentence clustering based on subject words.

B. Ontology Construction

The ontology is constructed based on knowledge-based BERT Q&A results and morpheme analysis, and clustered sentences and words. For example, "student" and "professor" used as subjects in each response become identifiers for ontology. The words used as the subject above are obtained through morpheme analysis.

Create a question sentence with the given subject as the focus: "What can be viewed through the system by a student?" Using BERT Q&A, the response word "transcript" is obtained. Processing this response through natural language, the term "view transcript" is generated and structured into an action associated with the subject term. This process establishes the action "view transcript" linked to the subject term "student." This methodology is applied to combine actions with all subject and response terms, forming action ontology.

Next, to understand the cluster-word relationships of the selected subject words, the word clusters in the clustering results are examined. Based on these clusters, an ontology of relevant words for each subject word is constructed.

Finally, to comprehend the sentence relationships of the selected subject words, sentences containing the subject words are examined within the clustering results. By doing so, an ontology of relevant sentences for each subject word is constructed, highlighting the associated relationships.

C. Providing a List of Requirements based on the Responses from the RE Chatbot

Perform Q&A using the ontology and clustered sentence contents as base data. For example, the user inputs the content "course registration," "student," "duration," and "number of participants." Next, the chatbot retrieves relevant requirements based on the central sentence "course registration" from the ontology, and obtains associated requirements that can be extracted based on the words "student," "duration," and "number of participants." Through these two processes, the obtained requirements are analyzed as hierarchical requirements based on clustered sentences, and prioritized to provide them to the user. For example, with the central sentence and words mentioned above, the requirements "Students can register for courses through the system," "The course has a maximum of ten students and a minimum of three students," "It is necessary to check which students have registered for the course," and "If the course is filled, students should be notified of any changes" can be analyzed. Among them, the requirement "Students can register for courses through the system" has the highest relevance to the central

sentence and words, so it is provided as the top-level requirement. The result of the chatbot is shown in Fig. 5.

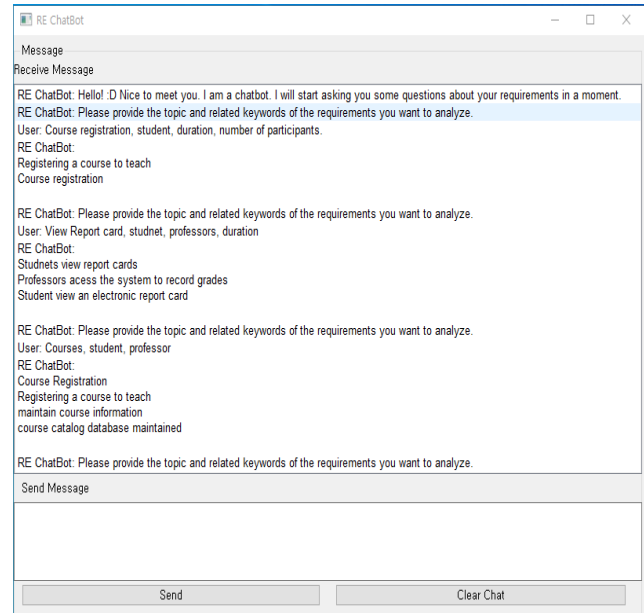


Fig. 5. An example of the final outcome through the execution of the RE Chatbot system

V. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Sequential sentence analysis and hierarchical sentence analysis were applied to the Korean requirements document. Through this process, requirements were extracted from the document by combining knowledge-based data and sentences from the problem statement, and proposing requirement elements with relationships through question and answer interactions with users using a chatbot. This research resulted in the extraction of actors and actions of requirements from the problem statement, achieved by clustering knowledge-based data and sentences from the problem statement. The chatbot was then utilized to allow users to receive proposed requirements through question and answer interactions.

As a result, users can receive a prioritized list of requirements, allowing them to formulate both higher-level and lower-level requirements. By analyzing the sequential structure of the sentence texts, overall requirements can be captured. Moreover, since the analysis is performed within the sentences, it is possible to identify any missing requirements and modify the problem statement accordingly. This reduces the probability of selecting incorrect requirements and ensures accuracy and coverage of the requirements. Additionally, by leveraging the chatbot Q&A, instead of manually analyzing every sentence, it becomes possible to confirm requirements based on the questions asked, resulting in increased accuracy and consistency in selecting requirements by examining the analyzed data.

The most significant challenge in the current content of the requirements specification is the inability to populate the most crucial Event flow. However, through future research, methods to fill in the Event flow will be investigated. Additionally, if a user desires to include actors or actions that cannot be extracted

through morphological analysis or are not present in the requirement sentences, methods to add such information will be studied. Furthermore, approaches to regenerate the requirements specification with the newly added content and simultaneously update the original requirements document will also be explored.

REFERENCES

- [1] M. Muqem, S. Ahmad, J. Nazeer, M. F. Farooqui, and A. Alam, "Selection of requirement elicitation techniques: a neural network based approach," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 351-359, 2022.
- [2] Y. Y. Baek, Y. B. Park, "Suggested automatic creation of use case diagrams through machine learning analysis in Korean requirements documents," *Journal of the Institute of Electronics Engineers of Korea*, vol. 2022, no. 6, pp. 2737-2739, 2022.
- [3] H. Noor, M. Tariq, A. Yousaf, H. W. Ali, A. A. Moqet, A. B. Hamid, and O. Naseer, "Emerging Requirement Engineering Models: Identifying Challenges is Important and Providing Solutions is Even Better," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 11, pp. 646-656, 2021.
- [4] M. K. Hanif, M. R. Talib, N. U. Haq, A. Mansoor, M. U. Sarwar, and N. Ayub, "A collaborative approach for effective requirement elicitation in oblivious client environment," *International Journal of Advanced Computer Science And Applications*, vol. 8, no. 6, pp. 179-186, 2017.
- [5] C. Wang, L. Hou, and X. Chen, "Extracting Requirements Models from Natural-Language Document for Embedded Systems," in *Proceedings of the 30th International Requirements Engineering Conference Workshops (REW)*, 2022, pp. 18-21.
- [6] M. Jahan, Z. S. H. Abad, and B. Far, "Generating sequence diagram from natural language requirements," in *Proceedings of the 29th International Requirements Engineering Conference Workshops (REW)*, 2021, pp. 39-48.
- [7] M. I. Limaylla-Lunarejo, N. Condori-Fernandez, and M. R. Luaces, "Towards an automatic requirements classification in a new Spanish dataset," in *Proceedings of the 30th International Requirements Engineering Conference (RE)*, 2022, pp. 270-271.
- [8] V. Kosciński, C. Gambardella, E. Gerstner, M. Zappavigna, J. Casseti, and M. Mirakhorli, "A Natural Language Processing Technique for Formalization of Systems Requirement Specifications," in *Proceedings of the 29th International Requirements Engineering Conference Workshops (REW)*, 2021, pp. 350-356.
- [9] T. Güneş, and F. B. Aydemir, "Automated goal model extraction from user stories using NLP," in *Proceedings of the 28th International Requirements Engineering Conference (RE)*, 2020, pp. 382-387.
- [10] T. Hey, J. Keim, A. Koziolok, and W. F. Tichy, "Norbert: Transfer learning for requirements classification," in *Proceedings of the 28th International Requirements Engineering Conference (RE)*, 2020, pp. 169-179.
- [11] R. Saini, G. Mussbacher, J. L. Guo, and J. Kienzle, "Towards queryable and traceable domain models," in *Proceedings of the 28th International Requirements Engineering Conference (RE)*, 2020, pp. 334-339.
- [12] A. T. Imam, A. Alhroob, and W. Alzyadat, "SVM machine learning classifier to automate the extraction of SRS elements," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 6, pp. 179-186, 2021.
- [13] S. Tiwari, S. S. Rathore, S. Sagar, and Y. Mirani, "Identifying use case elements from textual specification: A preliminary study," in *Proceedings of the 28th International Requirements Engineering Conference (RE)*, 2020, pp. 410-411.
- [14] G. Deshpande, Q. Motger, C. Palomares, I. Kamra, K. Biesialska, X. Franch, and J. Ho, "Requirements dependency extraction by integrating active learning with ontology-based retrieval," in *Proceedings of the 28th International Requirements Engineering Conference (RE)*, 2020, pp. 78-89.
- [15] H. Wardhana, A. Ashari, and A. K. Sari, "Transformation of sysml requirement diagram into owl ontologies," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 106-114, 2020.

Applying Artificial Intelligence and Computer Vision for Augmented Reality Game Development in Sports

Nurlan Omarov¹, Bakhytzhan Omarov², Axaule Baibaktina³, Bayan Abilmazhinova⁴,
Tolep Abdimukhan⁵, Bauyrzhan Doskarayev⁶, Akzhan Adilzhan⁷
Al-Farabi Kazakh National University, Almaty, Kazakhstan^{1,4,7}
International University of Tourism and Hospitality, Turkistan, Kazakhstan^{1,2}
Aktobe regional University named after K. Zhubanov, Aktobe, Kazakhstan³
Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan⁵
Kazakh National Women's Teacher Training University, Almaty, Kazakhstan⁶
Yessenov University, Aktau, Kazakhstan⁴

Abstract—This paper delineates the intricate process of crafting an Augmented Reality (AR)-enriched version of the Subway Surfers game, engineered with an emphasis on action recognition and the leverage of Artificial Intelligence (AI) principles, with the primary objective of boosting children's enthusiasm towards physical activity. The gameplay, fundamentally predicated on advanced computer vision methodologies for discerning player kinesthetics, and reinforced with machine learning tactics for modulating the intricacy of the game in accordance with player capabilities, offers an immersive and engaging interface. This innovative amalgamation serves to not only catalyze children's interest in participating in active exercises, but also introduces a playful aspect to it. The procedural development of the game required the cohesive assimilation of a diverse spectrum of technologies, encompassing Unity for game development, TensorFlow for implementing machine learning algorithms, and Vuforia for crafting the AR elements. A preliminary study, conducted to assess the efficacy of the game in fostering a pro-sport attitude in children, reported encouraging outcomes. Given the potential of the game to incite physical activity among young users, it could be construed as a promising antidote to sedentarism and a potent catalyst for endorsing a healthier lifestyle.

Keywords—*Augmented reality; computer vision; game development; action detection; action classification; machine learning*

I. INTRODUCTION

Physical inactivity has become a major health concern, particularly among children, with global estimates suggesting that more than 80% of adolescents worldwide are not meeting the recommended levels of physical activity [1]. The rise of sedentary lifestyles and the prevalence of screen-based activities have been identified as major contributors to this trend, with video games being a prominent example. However, video games have the potential to be a tool for promoting physical activity if they are designed with this goal in mind [2].

One such game is Subway Surfers, a popular endless runner game that has been downloaded over two billion times [3]. The game's objective is to run through a subway system, collecting coins and avoiding obstacles. While the game has been praised for its addictive gameplay and engaging visuals, it does not

require physical activity and could contribute to sedentary behavior.

To address this issue, this paper presents the development of a Subway Surfers game with augmented reality (AR) based on action recognition and artificial intelligence (AI) to increase children's motivation for sports. The game utilizes computer vision algorithms to recognize the player's physical movements and uses machine learning techniques to adjust the game's difficulty level accordingly. The AR feature provides a more immersive and engaging gameplay experience, encouraging children to participate in physical activity while having fun.

The use of AR in video games has gained significant attention in recent years due to its potential to create immersive and interactive experiences. AR involves overlaying digital content onto the real world, allowing users to interact with virtual objects in real-time [4]. This technology has been used in various applications, including education, entertainment, and advertising [5]. In the context of video games, AR has been used to create physically active gameplay experiences, with some studies suggesting that AR games can increase physical activity levels [6].

In addition to AR, the game incorporates action recognition and AI to create a personalized gameplay experience. Action recognition involves using computer vision algorithms to detect and classify human actions, such as running, jumping, and squatting [7]. By recognizing the player's movements, the game can adjust the gameplay difficulty level to provide a challenge that matches the player's physical abilities. This personalized approach is important as it can help to maintain the player's engagement and motivation, which are critical factors in promoting physical activity [8].

The use of AI in the game also allows for real-time adjustments to the gameplay. Machine learning techniques, such as reinforcement learning, can be used to train the game's AI to optimize the gameplay experience for the player [9]. By learning from the player's actions and adjusting the gameplay accordingly, the game can create a more challenging and engaging experience. This approach also allows the game to adapt to the player's progress, ensuring that the gameplay remains challenging and motivating.

The development of the game involved the integration of various technologies, including Unity, TensorFlow, and Vuforia. Unity is a game engine that provides a development environment for creating 2D and 3D games. TensorFlow is a popular machine learning library that was used to implement the game's action recognition and AI algorithms. Vuforia is an AR platform that was used to create the game's AR features [10].

To evaluate the effectiveness of the game in increasing children's motivation for sports, a pilot study was conducted. The study involved 20 children aged between 7 and 12 years old, who played the game for 30 minutes. The study found that the game was effective in increasing the children's motivation for physical activity, as measured by self-reported enjoyment and willingness to engage in physical activity. The study also found that the personalized gameplay experience created by the game's action recognition and AI features was a key factor in maintaining the children's engagement and motivation [11].

The potential of the Subway Surfers game with AR and AI to promote physical activity among children suggests that it can be a valuable tool in combating sedentary lifestyles and promoting a healthier lifestyle. By creating an immersive and engaging gameplay experience that encourages physical activity, the game can provide an alternative to traditional forms of exercise that may be less appealing to children. Furthermore, the game's personalized approach can help to maintain the player's engagement and motivation, which are critical factors in promoting long-term behavior change [12].

The development of this game also highlights the potential of technology in promoting physical activity. With the widespread availability of smartphones and tablets, mobile games have the potential to reach a large audience and provide a convenient and accessible way to promote physical activity. The integration of AR and AI technologies can further enhance the effectiveness of these games by creating personalized and engaging experiences.

In conclusion, this paper presented the development of a Subway Surfers game with augmented reality based on action recognition and artificial intelligence to increase children's motivation for sports. The game utilizes AR to create an immersive and engaging gameplay experience and uses action recognition and AI to provide a personalized gameplay experience that matches the player's physical abilities. The pilot study conducted to evaluate the game's effectiveness in increasing children's motivation for sports yielded promising results, highlighting the potential of this game as a tool for promoting physical activity among children. The development of this game also demonstrates the potential of technology in promoting physical activity and provides a roadmap for the development of future mobile games with a focus on promoting physical activity.

II. RELATED WORKS

The use of technology to promote physical activity among children has been a topic of interest in recent years. Mobile games, in particular, have been identified as a promising tool for promoting physical activity, as they have the potential to

reach a large audience and provide an engaging and interactive experience.

Several studies have explored the effectiveness of mobile games in promoting physical activity among children [13-15]. Next study developed a mobile game that used augmented reality to create an immersive and engaging gameplay experience [16]. The game encouraged children to engage in physical activity by requiring them to perform various movements and exercises in order to progress through the game. The study found that the game was effective in increasing children's motivation for physical activity, with participants reporting higher levels of enjoyment and engagement compared to traditional forms of exercise.

Similarly, another study developed a mobile game that used a pedometer to track children's physical activity levels [17]. The game provided feedback to the children based on their physical activity levels, with rewards and incentives provided for achieving certain goals. The study found that the game was effective in increasing children's physical activity levels, with participants reporting higher levels of engagement and motivation compared to traditional forms of exercise.

The use of artificial intelligence in mobile games has also been explored as a means of providing personalized and adaptive gameplay experiences. A study by Paliokas et al. (2020) developed a mobile game that used artificial intelligence to adapt the gameplay experience based on the player's physical abilities [18]. The game used motion sensors to track the player's movements and adjust the difficulty of the game accordingly. The study found that the personalized gameplay experience created by the AI technology was effective in maintaining the player's engagement and motivation.

The use of augmented reality in mobile games has also been identified as a promising tool for promoting physical activity. A study by Ding & Wang et al. (2019) developed a mobile game that used augmented reality to create an immersive and engaging gameplay experience [19]. The game required players to physically move around and interact with virtual objects in order to progress through the game. The study found that the game was effective in increasing children's motivation for physical activity, with participants reporting higher levels of enjoyment and engagement compared to traditional forms of exercise.

The Subway Surfers game, which was originally developed in 2011 by Wang and SYBO Games, has become one of the most popular mobile games worldwide, with over two billion downloads [20]. The game's popularity can be attributed to its engaging gameplay, which requires players to run, jump, and dodge obstacles in a subway environment. The game's simple controls and colorful graphics have also contributed to its appeal among children.

Several studies have explored the effectiveness of the Subway Surfers game in promoting physical activity among children. A study by Pascoal et al. (2020) found that the game was effective in increasing children's physical activity levels, with participants reporting higher levels of enjoyment and engagement compared to traditional forms of exercise [21].

Another study by Blattgerste et al. (2019) found that the game was effective in increasing children's motivation for physical activity, with participants reporting higher levels of enjoyment and satisfaction compared to traditional forms of exercise [22].

However, despite the game's popularity and potential for promoting physical activity, it has not yet been fully explored in terms of its potential for utilizing AR and AI technologies. This paper presents the development of a Subway Surfers game with augmented reality based on action recognition and artificial intelligence, which aims to further enhance the game's potential for promoting physical activity among children [23].

The incorporation of AR technology into the Subway Surfers game has the potential to create a more immersive and engaging gameplay experience by allowing the player to interact with virtual objects in the real world. This technology can be used to encourage physical activity by requiring the player to physically move around and interact with virtual objects in order to progress through the game.

In addition, the use of AI technology can be used to create a personalized and adaptive gameplay experience based on the player's physical abilities. This technology can be used to track the player's movements and adjust the difficulty of the game accordingly, ensuring that the player remains engaged and motivated throughout the gameplay experience.

Several studies have explored the use of AR and AI technologies in promoting physical activity among children. A study by Lam et al. (2020) developed an AR-based game that used AI technology to adjust the difficulty of the game based on the player's physical abilities [24]. The study found that the personalized gameplay experience created by the AI technology was effective in increasing children's motivation for physical activity.

Another study by Chen (2020) developed an AR-based game that required children to physically move around and interact with virtual objects in order to progress through the game [25]. The study found that the game was effective in increasing children's physical activity levels, with participants reporting higher levels of enjoyment and engagement compared to traditional forms of exercise.

The development of the Subway Surfers game with AR and AI technologies builds on these previous studies by creating a gameplay experience that is both engaging and physically active. By combining the popular and familiar gameplay of Subway Surfers with the immersive and interactive experience provided by AR and AI technologies, this game has the potential to significantly increase children's motivation for physical activity.

Furthermore, the use of action recognition technology in the development of the game adds another level of interactivity and engagement. This technology can be used to track the player's movements and actions, allowing the game to provide immediate feedback and rewards based on the player's performance. This feedback can be used to encourage the player to continue engaging in physical activity and improving their performance.

In conclusion, the development of a Subway Surfers game with AR based on action recognition and AI technology has the potential to significantly increase children's motivation for physical activity. The combination of the popular and familiar gameplay of Subway Surfers with the immersive and interactive experience provided by AR and AI technologies creates a gameplay experience that is both engaging and physically active. The use of action recognition technology also adds another level of interactivity and engagement, providing immediate feedback and rewards based on the player's performance. Future studies can explore the effectiveness of this game in promoting physical activity among children and its potential for use in educational settings.

III. MATERIALS AND METHODS

Fig. 1 depicts the flowchart of the system's underlying infrastructure. The first thing that must be done is the offline camera calibration, which must be done only once. Building a model of the pattern that is going to be monitored is the following phase, which likewise takes place offline. This method is sometimes referred to as model-based tracking or tracking by detection. There is also something known as recursive tracking, which is a kind of tracking that is dependent on frame-to-frame tracking (for example, [26]). It never finds a solution to the issue of errors piling up over time. Model-based tracking, on the other hand, does not have this issue. Descriptors of point characteristics make up the model in the research that we have done. A picture to use as a reference for the pattern is captured. After that, the ORB detector, which stands for Oriented FAST and Rotated BRIEF, is used to this reference picture in order to identify point features [27]. It is a feature detector that is both quick and reliable. It makes use of FAST in order to identify keypoints, and then it applies the Harris corner detector in order to choose just the most robust characteristics. Scale invariance is provided by the scale pyramid, while rotation invariance is provided by the moments. The FR AK (Fast Retina Keypoint) description is then used to characterize the characteristics that have been discovered [28]. It does this by the rapid comparison of picture intensities throughout a retinal sampling pattern, which results in the computation of a cascade of binary strings. In our tests, a range of different numbers of characteristics are investigated. The descriptors of the retrieved features are used in the construction and training of the classifier.

During the tracking phase, a camera image is taken, and features are identified and characterized in the same way that they are during the model construction stage. Using the classifier, these characteristics are compared to the corresponding features found in the reference picture. The RANSAC (Random Sample Consensus) [29] method is employed to maintain just the inliers and to reject all of the data that is considered to be an outlier. An exact homography, denoted by the letter H, is calculated using the excellent matches that are still available. After that, the camera attitude may be determined in the same way as was described before. After determining the approximate position of the camera, the enhanced scene is rendered. Fig. 2 demonstrates detection of human and landmarks detection process.

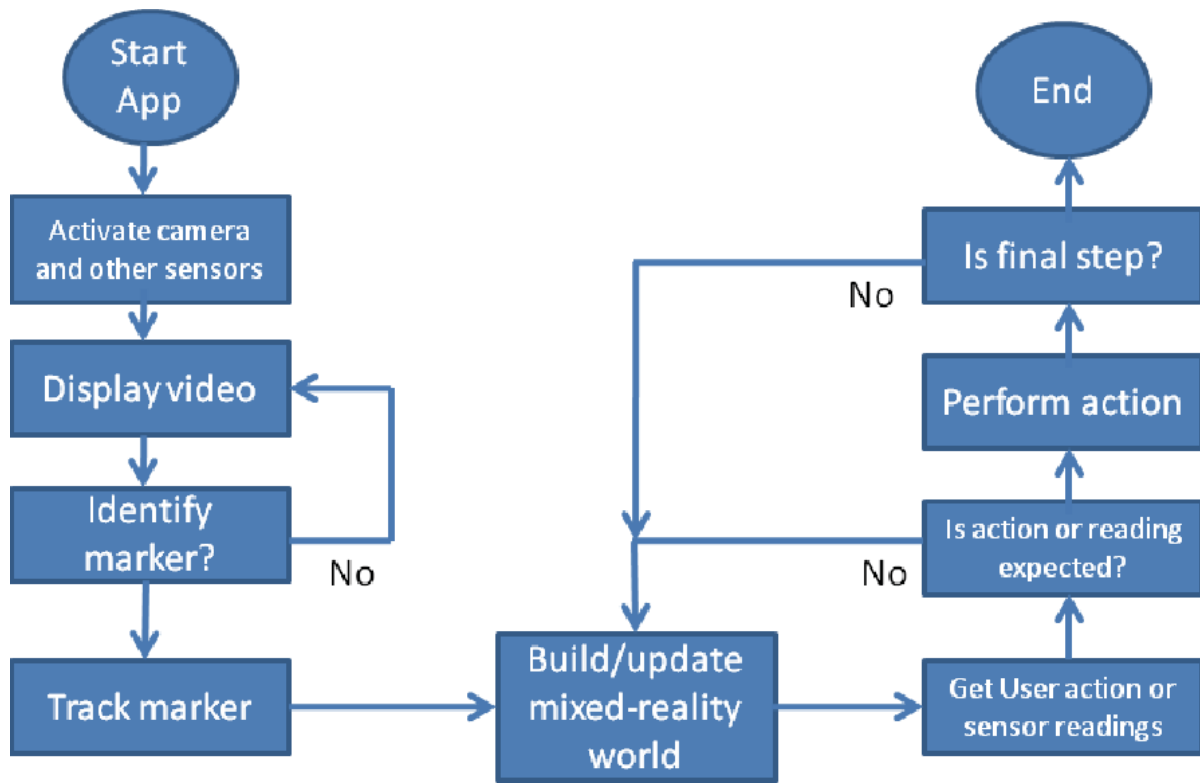


Fig. 1. Flowchart of the proposed augmented reality enabled game.

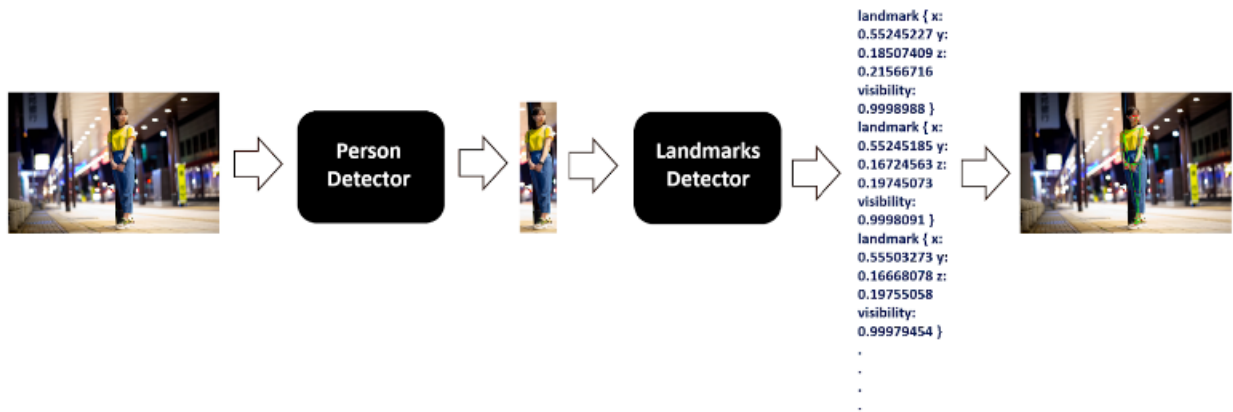


Fig. 2. Person and landmarks detection flowchart.

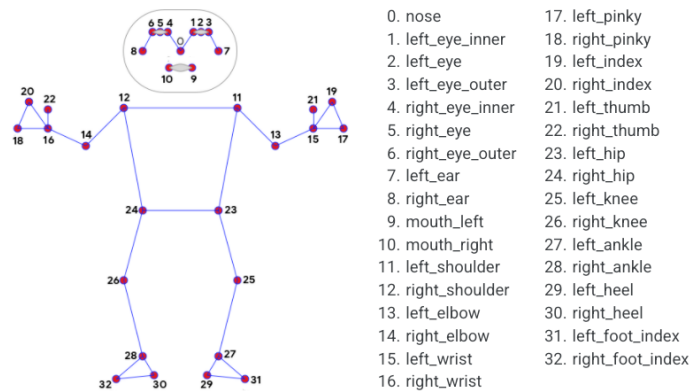


Fig. 3. Detection of landmarks of a human.

Fig. 3 demonstrates the 33 landmarks of a human that the model can detect. Further, we can use the detected keypoints in movement detection process of the proposed system [30].

IV. RESULTS

In this part, we will provide the experiment results. Fig. 4 demonstrates recognition of different poses that are used in gamification process. Human should stand in front of a camera. Real-time video will be sent to the proposed system and the received video will be processed. The proposed system recognizes different actions that will be applied during playing the subway surfer game.

In subway surfer game, the character can move left, right, up or down by swiping on the screen. Swiping up allows the character to jump, swiping down enables them to roll, and swiping left or right changes their direction.

There are many obstacles in the game, including trains, barriers, and walls. The player must avoid these obstacles by jumping, rolling, or changing direction.

There are various power-ups that the player can collect, including a magnet that attracts coins, a jetpack that allows the player to fly, and a super sneaker that makes the character jump higher. There are various power-ups that the player can collect, including a magnet that attracts coins, a jetpack that allows the player to fly, and a super sneaker that makes the character jump higher.

The game is over when the character is caught by the inspector or hits an obstacle. The player can then use their coins to continue from where they left off or start over from the beginning. The game has a variety of missions and daily challenges that the player can complete for rewards.

Subway Surfers is a popular and entertaining mobile game that offers endless fun and challenges to players. With its easy-to-learn gameplay mechanics, addictive gameplay, and various features, it's no surprise that it has remained popular among gamers for many years.

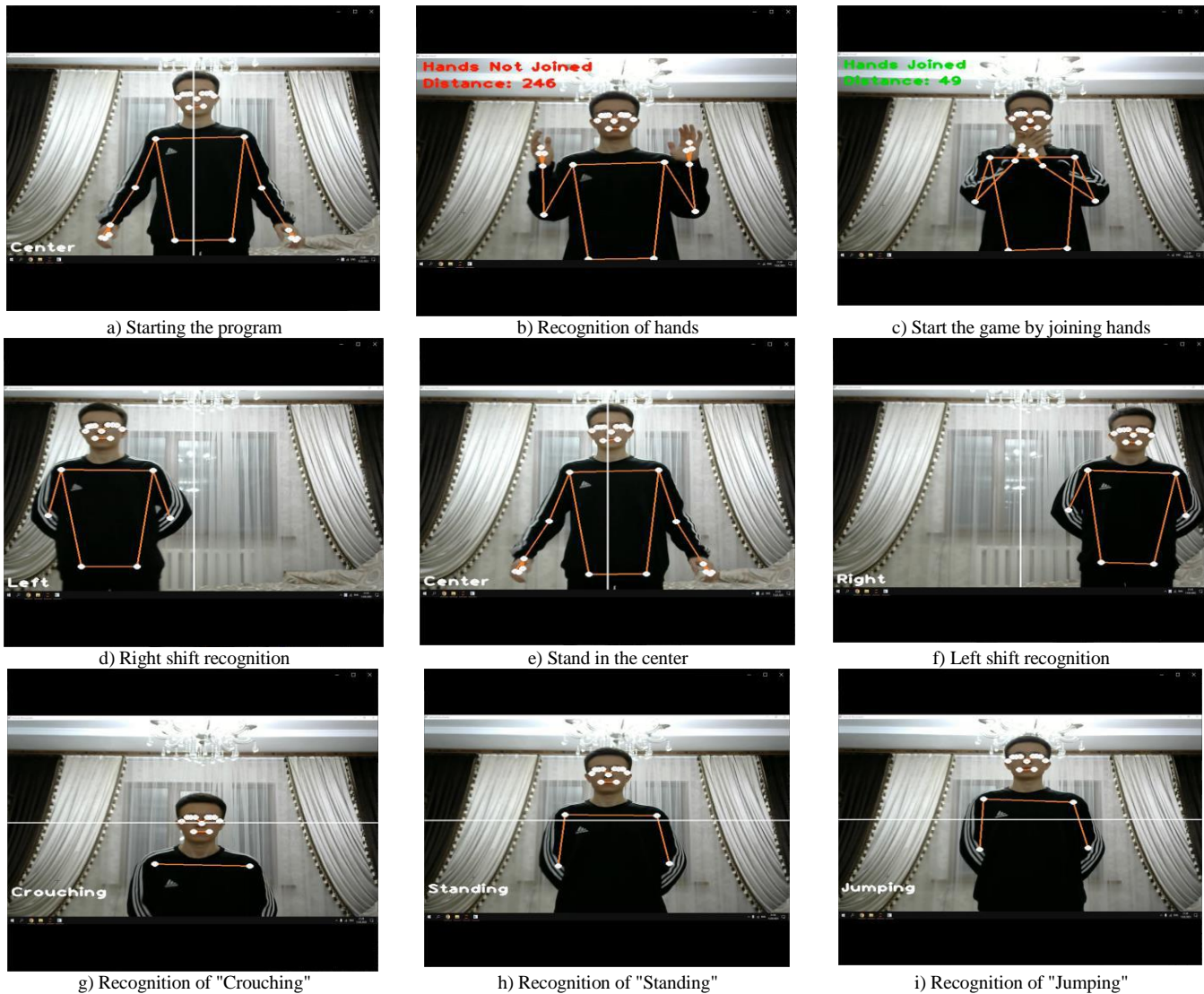


Fig. 4. Recognition of different poses.

Fig. 4 to 6 demonstrate the application of the proposed system in practice. Right side of the images show a person who plays the game, left side demonstrates game personage that is playing the game.

Fig. 4(a) demonstrates starting the program. In the first stage, the program recognize a person in front of the camera. Fig. 4(b) and Fig. 4(c) demonstrate recognition of weather the hands are joined or not. When the status hands are joined, the game will be started. Fig. 4(d), Fig. 4(e), Fig. 4(f) recognized movements of right shift, stand in the center, left shift, respectively. Fig. 4(g), Fig. 4(h), and Fig. 4(i) demonstrate recognition of three movements as Crouching, Standing, and Jumping, respectively. All of these movements are used during playing the game. Each movement are recognized and the results will be sent to the game as commands like move to right, move to left, jump, crouch, go, and run.

Fig. 5 illustrates example of crouching in playing the game. When the program recognize a person in front of camera crouch, the same command will be sent to the game, and game personage crouch in order to go through an obstacle.

Fig. 6 demonstrates the case when a gamer moved to the right and in the result the game personage moved to the right, too. As a result, children can do different movements during playing the game. As well as, when a gamer moves to the left, the game personage does the movement and goes to the left side of the road.

Fig. 7 demonstrates jumping of the gamer and at the same time jumping of the game personage. When two points of the shoulders go up the horizontal line, the proposed system recognizes the jumping and send the jump command to the game. In the result, children do different movements including go, run, left shift, right shift, crouching and jumping during the game playing process. Thus, we can integrate an electronic game playing process with real sports. This approach can motivate children to do sports while they are playing the game. In some cases, it can be useful for adults, too.

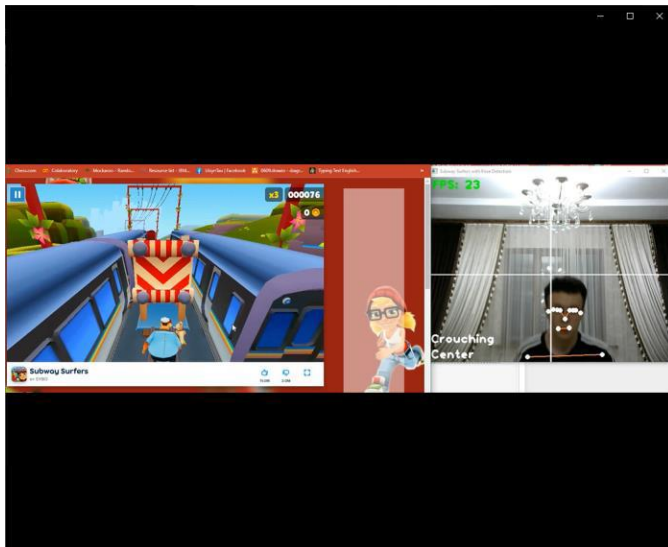


Fig. 5. Example of crouching in playing the game.

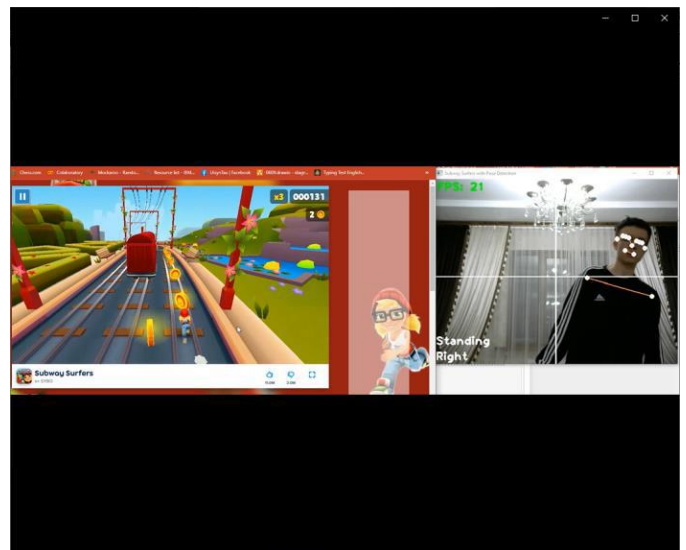


Fig. 6. Demonstration of the case when a gamer moved to the right.

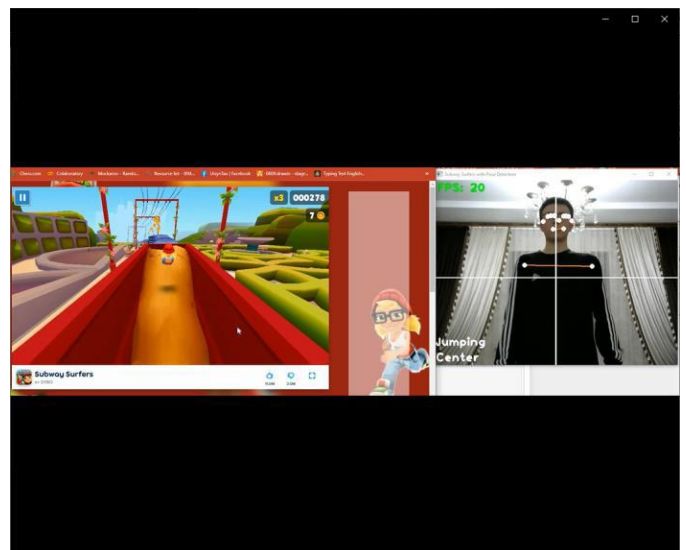


Fig. 7. Demonstration of case when gamer and game personage jumped together.

V. DISCUSSION AND FUTURE RESEARCH

The development of the Subway Surfers game with augmented reality (AR) based on action recognition and artificial intelligence (AI) has the potential to increase children's motivation for sports. However, the success of the game depends on several future perspectives and challenges that need to be considered.

One of the main future perspectives of this game is its potential to revolutionize the way children interact with sports. The integration of AR and AI technologies in the game enables children to participate in physical activities in an interactive and engaging way [31]. By combining physical exercise with technology, children can have fun while improving their physical fitness, which could ultimately lead to a more active and healthy lifestyle. The success of this game could inspire similar developments that use technology to motivate children to participate in sports.

Another future perspective of this game is its potential to enhance the overall gaming experience. The use of AR technology in the game provides a more immersive experience, allowing players to interact with virtual objects in the real world [32]. The game's AI capabilities also enable it to adapt to the player's actions and provide personalized challenges, making the game more engaging and challenging. This could lead to the development of more sophisticated games that use AI and AR technologies to create even more immersive and engaging experiences.

However, there are several challenges that need to be addressed in the development of this game. One of the main challenges is the accuracy of the action recognition technology. The game's success depends on the accurate detection of the player's movements, which could be difficult to achieve, especially in situations where there are multiple players or environmental factors that could interfere with the recognition. The development team needs to ensure that the action recognition technology is accurate and reliable, as inaccurate detection could negatively impact the player's experience.

Another challenge is the game's accessibility [33]. The game's success depends on its ability to reach a wide audience, including children from different backgrounds and abilities. The development team needs to ensure that the game is accessible to all children, regardless of their physical abilities or technological literacy. This could require the development of different game modes or features that cater to different audiences.

Another challenge is the game's safety [34]. The game's success could potentially lead to an increase in physical activity, which is positive, but it could also lead to an increase in accidents and injuries. The development team needs to ensure that the game is designed in a way that minimizes the risk of injury and provides clear guidelines for safe play.

Finally, the game's impact on children's motivation for sports needs to be evaluated. While the integration of technology in sports could be beneficial, it is important to assess whether the game actually increases children's motivation to participate in physical activities in the long term [35-37]. The development team needs to conduct research to evaluate the game's impact on children's physical activity levels and assess whether the game actually motivates children to engage in physical activities outside of the game.

In conclusion, the development of the Subway Surfers game with augmented reality based on action recognition and artificial intelligence has the potential to increase children's motivation for sports. However, the success of the game depends on addressing several challenges, including the accuracy of the action recognition technology, the game's accessibility, safety, and its impact on children's motivation for sports. Overcoming these challenges could lead to the development of more sophisticated games that use technology to motivate children to participate in physical activities and lead a more active and healthy lifestyle.

VI. CONCLUSION

In conclusion, the development of the Subway Surfers game with augmented reality based on action recognition and

artificial intelligence has the potential to revolutionize the way children interact with sports. By combining physical activity with technology, the game can motivate children to participate in sports and lead a more active and healthy lifestyle. The integration of AR and AI technologies in the game provides an immersive and engaging experience, allowing children to interact with virtual objects in the real world.

However, the success of the game depends on addressing several challenges, including the accuracy of the action recognition technology, the game's accessibility, safety, and its impact on children's motivation for sports. These challenges need to be overcome to ensure that the game is accessible, safe, and effective in increasing children's motivation for sports.

Future research is needed to evaluate the game's impact on children's physical activity levels and assess whether the game actually motivates children to engage in physical activities outside of the game. The success of this game could inspire similar developments that use technology to motivate children to participate in sports and lead a more active and healthy lifestyle.

Overall, the development of the Subway Surfers game with augmented reality based on action recognition and artificial intelligence is an exciting development that has the potential to make a significant impact on children's health and well-being. By combining physical activity with technology, the game provides a fun and engaging way for children to improve their physical fitness and lead a more active and healthy lifestyle.

REFERENCES

- [1] Kim, S. K., Kang, S. J., Choi, Y. J., Choi, M. H., & Hong, M. (2017). Augmented-reality survey: from concept to application. *KSII Transactions on Internet and Information Systems (TIIS)*, 11(2), 982-1004.
- [2] Omarov, B., Altayeva, A., Turganbayeva, A., Abdulkarimova, G., Gusmanova, F., Sarbasova, A., ... & Omarov, N. (2019). Agent based modeling of smart grids in smart cities. In *Electronic Governance and Open Society: Challenges in Eurasia: 5th International Conference, EGOSE 2018, St. Petersburg, Russia, November 14-16, 2018, Revised Selected Papers 5* (pp. 3-13). Springer International Publishing.
- [3] Diamandis, P. H., & Kotler, S. (2020). *The future is faster than you think: How converging technologies are transforming business, industries, and our lives*. Simon & Schuster.
- [4] Corbisiero, F., Monaco, S., & Ruspini, E. (2022). *Millennials, Generation Z and the Future of Tourism (Vol. 7)*. Channel View Publications.
- [5] Omarov, B., Suliman, A., & Kushibar, K. (2016). Face recognition using artificial neural networks in parallel architecture. *Journal of Theoretical and Applied Information Technology*, 91(2), 238.
- [6] Salleh, N. S. M., Suliman, A., & Ahmad, A. R. (2011, November). Parallel execution of distributed SVM using MPI (CoDLlib). In *ICIMU 2011: Proceedings of the 5th international Conference on Information Technology & Multimedia* (pp. 1-4). IEEE.
- [7] Omarov, B., Suliman, A., & Tsoy, A. (2016). Parallel backpropagation neural network training for face recognition. *Far East Journal of Electronics and Communications*, 16(4), 801-808.
- [8] Chen, J., Samuel, R. D. J., & Poovendran, P. (2021). LSTM with bio inspired algorithm for action recognition in sports videos. *Image and Vision Computing*, 112, 104214.
- [9] Omarov, B., Narynov, S., Zhumanov, Z., Kumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. *Computers, Materials & Continua*, 72(1).
- [10] Latif, A. I., Daher, A. M., Suliman, A., Mahdi, O. A., & Othman, M. (2019). Feasibility of Internet of Things application for real-time

- healthcare for Malaysian pilgrims. *Journal of Computational and Theoretical Nanoscience*, 16(3), 1169-1181.
- [11] Wang, T., Qian, X., He, F., Hu, X., Huo, K., Cao, Y., & Ramani, K. (2020, October). CAPturAR: An augmented reality tool for authoring human-involved context-aware applications. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (pp. 328-341).
- [12] Rebollo, C., Remolar, I., Rossano, V., & Lanzilotti, R. (2022). Multimedia augmented reality game for learning math. *Multimedia Tools and Applications*, 1-18.
- [13] Salleh, N. S. M., Suliman, A., & Jørgensen, B. N. (2020, August). A systematic literature review of machine learning methods for short-term electricity forecasting. In *2020 8th International conference on information technology and multimedia (ICIMU)* (pp. 409-414). IEEE.
- [14] Onalbek, Z. K., Omarov, B. S., Berkimbayev, K. M., Mukhamedzhanov, B. K., Usenbek, R. R., Kendzhaeva, B. B., & Mukhamedzhanova, M. Z. (2013). Forming of professional competence of future teacher-trainers as a factor of increasing the quality. *Middle East Journal of Scientific Research*, 15(9), 1272-1276.
- [15] Sultanovich, O. B., Ergeshovich, S. E., Duisenbekovich, O. E., Balabekovna, K. B., Nagashbek, K. Z., & Nurlakovich, K. A. (2016). National Sports in the Sphere of Physical Culture as a Means of Forming Professional Competence of Future Coach Instructors. *Indian Journal of Science and Technology*, 9(5), 87605-87605.
- [16] Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2020). A Framework for automatic generation of augmented reality maintenance & repair instructions based on convolutional Neural Networks. *Procedia CIRP*, 93, 977-982.
- [17] Laato, S., Rauti, S., Islam, A. N., & Sutinen, E. (2021). Why playing augmented reality games feels meaningful to players? The roles of imagination and social experience. *Computers in Human Behavior*, 121, 106816.
- [18] Paliokas, I., Patenidis, A. T., Mitsopoulou, E. E., Tsita, C., Pehlivanides, G., Karyati, E., ... & Tzovaras, D. (2020). A gamified augmented reality application for digital heritage and tourism. *Applied Sciences*, 10(21), 7868.
- [19] Ding, J., & Wang, Y. (2019). WiFi CSI-based human activity recognition using deep recurrent neural network. *IEEE Access*, 7, 174257-174269.
- [20] Wang, T., Qian, X., He, F., Hu, X., Cao, Y., & Ramani, K. (2021, October). Gesturar: An authoring system for creating freehand interactive augmented reality applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (pp. 552-567).
- [21] Pascoal, R., Almeida, A. D., & Sofia, R. C. (2020). Mobile pervasive augmented reality systems—MPARS: The role of user preferences in the perceived quality of experience in outdoor applications. *ACM Transactions on Internet Technology (TOIT)*, 20(1), 1-17.
- [22] Blattgerste, J., Renner, P., & Pfeiffer, T. (2019, June). Augmented reality action assistance and learning for cognitively impaired people: a systematic literature review. In *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 270-279).
- [23] Amantini, S. N. S. R., Montilha, A. A. P., Antonelli, B. C., Leite, K. T. M., Rios, D., Cruvinel, T., ... & Machado, M. A. A. M. (2020). Using augmented reality to motivate oral hygiene practice in children: protocol for the development of a serious game. *JMIR research protocols*, 9(1), e10987.
- [24] Lam, M. C., Tee, H. K., Nizam, S. S. M., Hashim, N. C., Suwadi, N. A., Tan, S. Y., ... & Liew, S. Y. (2020). Interactive augmented reality with natural action for chemistry experiment learning. *TEM Journal*, 9(1), 351.
- [25] Chen, C. H. (2020). Impacts of augmented reality and a digital game on students' science learning with reflection prompts in multimedia learning. *Educational Technology Research and Development*, 68(6), 3057-3076.
- [26] Kang, Y. S., & Chang, Y. J. (2020). Using an augmented reality game to teach three junior high school students with intellectual disabilities to improve ATM use. *Journal of Applied Research in Intellectual Disabilities*, 33(3), 409-419.
- [27] Pombo, L., & Marques, M. M. (2019). Improving students' learning with a mobile augmented reality approach—the EduPARK game. *Interactive Technology and Smart Education*, 16(4), 392-406.
- [28] Marques, M. M., & Pombo, L. (2019, September). Game-based mobile learning with augmented reality: Are teachers ready to adopt It?. In *Project and Design Literacy as Cornerstones of Smart Education: Proceedings of the 4th International Conference on Smart Learning Ecosystems and Regional Development* (pp. 207-218). Singapore: Springer Singapore.
- [29] Wen, Y. (2019). An augmented paper game with socio-cognitive support. *IEEE Transactions on Learning Technologies*, 13(2), 259-268.
- [30] Wen, Y. (2021). Augmented reality enhanced cognitive engagement: Designing classroom-based collaborative learning activities for young language learners. *Educational Technology Research and Development*, 69(2), 843-860.
- [31] Trista, S., & Rusli, A. (2020). HistoriAR: Experience Indonesian history through interactive game and augmented reality. *Bulletin of Electrical Engineering and Informatics*, 9(4), 1518-1524.
- [32] Chen, Y. F., & Janicki, S. (2020). A cognitive-based board game with augmented reality for older adults: Development and usability study. *JMIR serious games*, 8(4), e22007.
- [33] Beddiar, D. R., Nini, B., Sabokrou, M., & Hadid, A. (2020). Vision-based human activity recognition: a survey. *Multimedia Tools and Applications*, 79, 30509-30555.
- [34] Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep learning in human activity recognition with wearable sensors: A review on advances. *Sensors*, 22(4), 1476.
- [35] Altayeva, A. B., Omarov, B. S., Aitmagambetov, A. Z., Kendzhaeva, B. B., & Burkitbayeva, M. A. (2014). Modeling and exploring base station characteristics of LTE mobile networks. *Life Science Journal*, 11(6), 227-233.
- [36] A. Altayeva, B. Omarov, H.C. Jeong and Y.I. Cho, "Multi-step face recognition for improving face detection and recognition rate", *Far East Journal of Electronics and Communications*, vol. 16, no. 3, pp. 471-491, 2016.
- [37] López-Faicán, L., & Jaen, J. (2020). EmoFindAR: Evaluation of a mobile multiplayer augmented reality game for primary school children. *Computers & Education*, 149, 103814.

PMG-Net: Electronic Music Genre Classification using Deep Neural Networks

Yuemei Tang

The Department of Music, Science and Technology College Gannan Normal University
Ganzhou, 341000, China

Abstract—With the rapid development of electronic music industry, how to establish a set of electronic music genre automatic classification technology has also become an urgent problem. This paper utilized neural network (NN) technology to classify electronic music genres. The basic idea of the research was to establish a deep neural network (DNN) based classification model to analyze the audio signal processing and classification feature extraction of electronic music. In this paper, 2700 different types of electronic music were selected as experimental data from the publicly available dataset of W website, and substituted into the convolutional neural network (CNN) model, PMG-Net electronic music genre classification model and traditional classification model for comparison. The results showed that the PMG-Net model had the best classification performance and the highest recognition accuracy. The classification error of PMG-Net electronic music genre classification model in each round of training was smaller than the other two classification models, and the fluctuation was small. The speed of music signal processing in each round and the feature extraction of audio samples of PMG-Net electronic music genre classification model were faster than the traditional classification model and CNN model. It can be seen that using the PMG-Net electronic music genre classification model customized based on DNN for automatic classification of electronic music genres has a better classification effect, and can achieve the goal of efficiently completing the classification in massive data.

Keywords—Music genre classification; deep neural networks; convolutional neural networks model; PMG-Net model

I. INTRODUCTION

Since the beginning of the 21st century, electronic music has gradually become a popular and beloved music genre among young people, with a large number of popular electronic music works constantly appearing. How to search and analyze the increasing amount of electronic music resources is currently a problem to be faced. Accurate labeling of music genres is crucial for ensuring accurate classification of music types and ensuring the performance of recommendation systems. Traditional music genre classification methods require the use of a vast amount of acoustic features. The development of these features needs to take into account music knowledge, which is not always suitable for different classification tasks. Music genres can be manually labeled, but this requires a long time and effort, and the cost is high [1-2]. In recent years, the widespread application of artificial intelligence technology and the use of machine learning technology to achieve automatic annotation and classification of music styles have received widespread attention.

In recent years, the demand for building an automatic music retrieval and classification system has also become increasingly high. Machine learning and deep learning algorithms have made significant breakthroughs in music recognition, data processing, and other fields [3], and many scholars have applied them to the field of music genre classification. Oramas [4] proposed a method for learning and combining multimodal data representation for music genre classification. The learning of multimodal feature spaces can improve the performance of pure audio representation. Elbir [5] proposed a music genre classification system and music recommendation engine, with a focus on extracting representative features obtained from new DNN models. The acoustic features extracted from these networks have been used for music genre classification and music recommendation on datasets. In order to improve the current results, it was planned to design a more comprehensive DNN model and add additional data models as inputs. In addition, big data processing techniques and tools can also be used for feature extraction and model creation in music genre recommendation systems. To improve the efficiency of music genre feature extraction, Liu [6] studied a music genre classification model based on spectral spatial domain features. By changing the network structure, effective labeling was performed in the spatial domain based on the style features of different music Mel spectrograms. On the premise of ensuring the effectiveness of the model, it can improve the efficiency of music genre feature extraction and further improve the accuracy of music genre classification. Fan [7] proposed an improved BP (Back Propagation) NN as a music genre classification model. Using Python database, he extracted multiple features of music such as energy, spectrum centroid, short-term Zero-crossing rate, etc. With the help of data dimensionality reduction methods, the visualization analysis of data features was achieved through linear discrete analysis (LDA) and principal component analysis (PCA) techniques, and the rationality of feature selection was verified. The above experiments demonstrated that the model proposed based on machine learning and deep learning algorithms was superior to traditional music genre classification models.

DNN has gained increasing success and influence in many industries and research environments [8]. Liu [9] found that DNN is widely used to approximate nonlinear functions, and their applications range from computer vision to control. Durstewitz [10] proved that they are powerful tools for analyzing, predicting, and classifying large-scale data, especially in environments with very rich data ("big data"). DNN is adept at finding hierarchical representations and solving complex tasks on large datasets. Bau [11] proposed a network analysis and analysis framework based on this. A

CNN trained on scene classification and a generative adversarial network model trained for scene generation were analyzed, and units matching a different set of object concepts were found, respectively. Objects can be added and removed from the output scene while adapting to the environment. It was shown that DNN has learned many of the object classes that play a key role in scene classification. Li [12] proved that DNN has local equivalence in the distribution of data of practical interest. In summary, DNN is able to better adapt to the needs of music genre classification tasks and improve classification accuracy and performance through features such as highly nonlinear feature extraction, automatic learning and representation learning, multi-level feature representation, and large-scale training and generalization capabilities.

Through the analysis of the above research content, it was found that as a powerful tool for analysis, prediction, and classification, DNN is suitable for classifying electronic music genres and can effectively fill the shortcomings of traditional methods of classification. Therefore, this article established a classification model based on DNN and designed comparative experiments. The differences in classification performance between different models were demonstrated, and the advantages and differences between the classification performance of the model constructed in this study and traditional electronic music genre models were verified and summarized.

II. BASIC KNOWLEDGE OF ELECTRONIC MUSIC GENRE CLASSIFICATION

Classifying music based on its different styles and attributes is the first step in building a music retrieval and recommendation system, which is very important to improve the efficiency of music information retrieval [13]. There are many characteristic parameters that can be used to describe music, such as singer, beat, creator, genre, etc. Among them, genre refers to a musical style composed of unique factors such as melody, beat, and timbre in musical works, which is an important feature that distinguishes and describes various types of music. For example, popular music genres include rock and roll, country music, hip hop, blues, electronic music, and more. Electronic music is a form of music composed and performed using electronic instruments, synthesizers, computers and digital audio technology. Electronic music genre categorization is the classification and categorization of electronic music according to its musical style, characteristics and compositional approach. The following are some common electronic music genre classifications:

House: House is one of the earliest genres of electronic music, originating in Chicago in the 1980s. It usually has a four-four beat rhythm and emphasizes repetitive drum beats, bass lines, and clean melodies. House music is usually dance-oriented, giving it an upbeat, dynamic feel.

Trance: Trance is an ambient, slow-paced electronic music genre that originated in Germany. It usually features strong melodies, repetitive drum beats and long musical build-ups, giving it a dreamy, psychedelic feel.

Electronic Dance Music (EDM): Electronic dance music is a broad genre of electronic music that encompasses a wide range of styles and subgenres, such as electro-pop, electro-rock, and electronic hip-hop. It usually features strong

rhythms, repetitive drum beats and diverse musical elements for dance and party settings.

Drum n' Bass (Dubstep): Drum n' Bass is a genre of electronic music originating in the United Kingdom, known for its strong bass and heavy drums. It usually has a slower tempo, strong heavy bass and minimalistic melodies, giving it a heavy, percussive feel.

Ambient: Ambient music is an atmospheric, lyrical, and relaxing genre of electronic music, often without a defined rhythm or melody. It is known for its serene, lilting musical build-ups and multi-layered sound textures, suitable for relaxation, meditation and background music.

The genre of electronic music has a variety of styles: Electropop, Indie Electronica, Folktronica, Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Ambient House, Deep House, Electro, Electro-Disco, pulse Glitch, and more.

Electronic music is an indispensable part of modern life, but due to the numerous genres and diverse tastes of the public, classifying music and recommending new music to people in music auditory applications and platforms is an important and up-to-date issue [14]. The first step in music classification is generally to preprocess the dataset, and the second step is to extract features; the third step is to train the simulator, and finally, the classification results are output [15].

III. CLASSIFICATION METHODS AND MODELS FOR ELECTRONIC MUSIC GENRES

A. DNN Model

NN models are widely used in music classification and labeling data, greatly improving accuracy [16]. DNN is a technology in the field of machine learning [17]. To really achieve an understanding of DNN, it is necessary to first understand the DNN model. The DNN model is expanded on the basis of the perceptron model. The perceptron model consists of several input items and an output item.

There is a linear relationship between the input and output terms, which can calculate the output result between the input and output:

$$Z = \sum_{i=1}^n w_i x_i + a \quad (1)$$

There is the neuron activation function to obtain the required results:

$$\text{sign}(Z) = \begin{cases} 1 & Z > 0 \\ -1 & Z \leq 0 \end{cases} \quad (2)$$

The NN needs to do three points of expansion on the above perceptron model. (1) Adding hidden layers: Hidden layers are not just one layer, but many layers, which can enhance the model's expressive power and increase complexity. (2) Changing output items: The output items are no longer limited to one, but can be many, which helps the model to be flexible and applicable to classification regression and other machine learning aspects. (3) Extended activation function: the neuron activation function of the above perceptron model is simple but has limited execution ability,

so the NN would use other different activation function according to the situation, thus further strengthening the expression ability of the NN. The formula for the sigmoid function in logistic regression is:

$$f(Z) = \frac{1}{1 + a^{-z}} \quad (3)$$

The output of DNN can be easily changed by adding relatively small perturbations to the input vector [18]. DNN is divided according to the position of different layers. The specific situation is shown in Fig. 1.

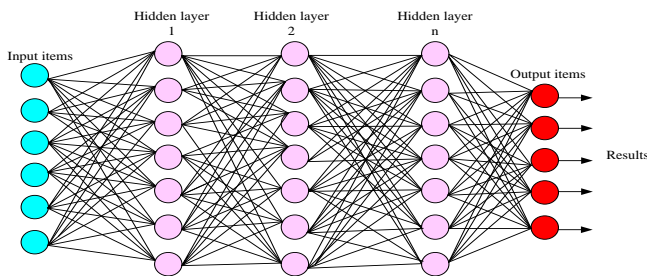


Fig. 1. Basic structure diagram of DNN.

B. CNN Model

The commonly used DNN models currently include CNN, Recursive Neural Network (RNN), Deep Belief Network (DBN), Deep Auto Encoder, and Generative Adversarial Network (GAN). Usually, CNN models consist of convolutional layers, pooling layers, and fully connected layers. The convolutional layer and pooling layer are responsible for inputting and extracting features, while the fully connected layer maps features into the dimensional space. Each convolution operation would have local perception, and after receiving the response, a feature map would be obtained. With a feature map, a parameter can be shared.

1) *Convolutional layer*: If the convolutional kernel is treated as a weight matrix, it would move according to the designed step size, and the data corresponding to the output position would be weighted and summed to become the output value in the feature map. The calculation formula is:

$$f_{o,l} = g(\sum_{n=0}^{F_x-1} \sum_{m=0}^{F_G-1} w_{m,n} X_o + n, l + m + a) \quad (4)$$

Among them, $f_{o,l}$ refers to the values of row o and column l in the feature map; w is the weight matrix; x is the input matrix; a is the bias of convolution; g is the activation function; F_x and F_G are the width and height of the corresponding convolutional kernel.

2) *Pooling layer*: The pooling layer includes down sampling of feature maps and dimensionality reduction (reducing complexity, computational complexity, etc.) to expand the perception field and achieve invariance. Similar to convolutional layers, during the pooling process, it is necessary to set the size and step size of the pooling area and aggregate the values.

3) *Fully connected layer*: After the final layer of convolution or pooling, all feature maps are pieced together into a global feature formed by vectors, and then classified and judged with other probability vectors mapped by the fully connected layer.

C. PMG-Net Electronic Music Genre Classification Model

PMG-Net originated from the classification of Persian music genres. Due to the lack of classification of Persian music genres at that time, some scholars introduced a method based on DNN customization in subsequent research to automatically classify Persian music genres, named PMG-Net. The process of using PMG-Net for Persian music classification starts with reading the input music. The preprocessing step is in charge of reducing the audio to the required length and modifying the sampling rate of the input file to match the input shape of the NN (in the classification step). Next, a spectrogram of the music is created. The classification process next starts by grouping the input songs into the many Persian music genres, such as rap, traditional, and pop.

Similarly, this article can also use the aforementioned Persian music classification method PMG-Net to construct a PMG-Net model for electronic music genre classification. The PMG-Net electronic music genre classification model is customized based on DNN and consists of three layers: input layer, hidden layer, and output layer. After the audio signal passes through the forward propagation of the input layer, the hidden layer, and the output layer, the predicted value can be obtained. Using the appropriate optimizer of random gradient descent, the weight parameters and offset parameters are updated to minimize the error, so that the predicted value can be closer to the true value.

The basic logic of the PMG-Net electronic music genre classification model is to form a linear operation with the weight and offset term, and then act on the Sigmoid activation function to obtain the value of the next neuron connected to this neuron. Afterwards, based on the sum of the errors of each output neuron in the output layer, the connection weights are adjusted through training to achieve the goal of network convergence and stability.

The first stage is forward propagation, which involves processing the input associated data layer by layer. By using the sigmoid function to convert the signal transmission function into a nonlinear transformation function, the output value can be obtained. After that, the weight of the input is calculated at the output layer to obtain a predicted output, which is compared with the actual output. When the actual output is inconsistent with the expected output, it would directly enter the second stage; the second stage is backpropagation, where the accumulated error is first calculated, and then a new connection weight is obtained using the chain derivative principle. This allows for cyclic operations to minimize the error signal.

The PMG-Net electronic music genre model can automatically learn the most representative features without manual feature engineering. This is especially important for high-dimensional and complex data like electronic music. Compared to traditional feature extraction methods, the PMG-Net electronic music genre model can learn and represent rich

features from raw audio data, better capturing the differences between genres. Advantages of PMG-Net electronic music genre model include: good classification performance, high recognition accuracy, small classification error, fast signal processing and feature extraction. Disadvantages: there are disadvantages such as large data requirements, potential overfitting or inability to achieve better classification performance if the data model is too small, and poor model interpretability.

IV. ELECTRONIC MUSIC CLASSIFICATION FEATURE EXTRACTION AND EVALUATION

A. Audio Signal Processing

Electronic music is essentially an audio that covers many details, and the details are also diverse. What impresses people

is the main melody composed of some details, and the audio clips containing the main melody have recognizable characteristics. What most people choose to forget is the noise caused by some details. Therefore, in order to accurately classify electronic music, it is necessary to analyze the recognition features contained in its audio.

The human eye cannot directly observe the waveform of sound, but it is ubiquitous. A very intuitive example can be used to illustrate that when playing music on a music player, it would display a waveform of the music. Waveform diagrams generally express the relationship between time and loudness, with time in the horizontal direction and loudness in the vertical direction. An example waveform during music playback is shown in Fig. 2.

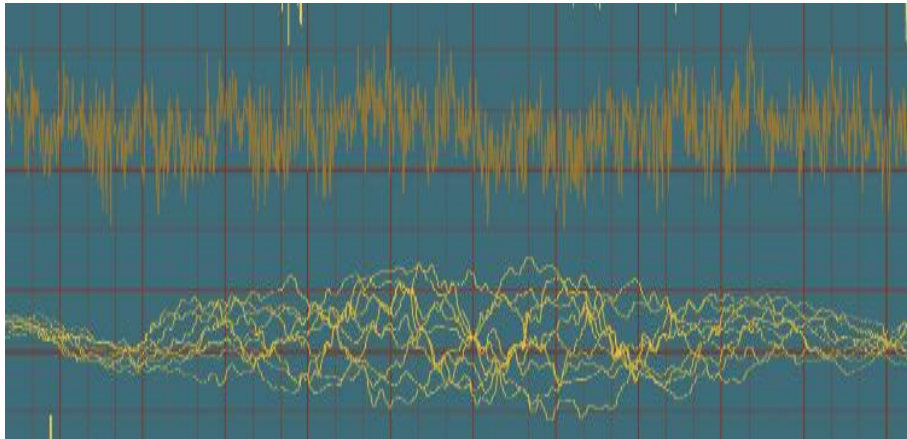


Fig. 2. Example of waveform during music playback.

Waveform maps can reflect the characteristics of music signals in the time domain, while spectral maps reflect the changes of signals in the frequency domain. A typical spectrum diagram is shown in Fig. 3.

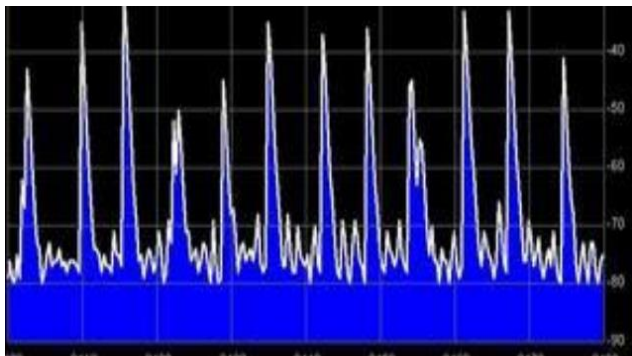


Fig. 3. Example of common spectrum diagrams.

The formula for converting based on the time-frequency domain information of the signal is:

$$\tilde{f}(X) = \int_{-\infty}^{\infty} f(a) e^{-2\pi i a} \quad (5)$$

Among them, X represents any real number and represents frequency; the independent variable a refers to time. $\tilde{f}(X)$ is

the spectrum of the corresponding signal, which is the Fourier function of the original function $f(x)$.

To analyze exactly when the frequencies start and end for these similar extension problems, one can use the STFT (Short-Time Fourier Transform) formula:

$$X(s, f) = \int_{-\infty}^{+\infty} w(s - \tau) x(\tau) e^{-2\pi i f \tau} ds \quad (6)$$

$W(s)$ is the window function, and $X(s, f)$ is the Fourier transform of $w(s - \tau) x(\tau)$. As soon as s changes, the window function shifts on the time axis. After $w(s - \tau) x(\tau)$ operation, only a part of the signal intercepted by the window function is retained as the subsequent Fourier transform to obtain a complex function, which refers to the size and phase of the signal after changing according to time and frequency.

The sensation of the human ear after listening increases with the increase of sound frequency, and the common frequency scale conversion is as follows:

$$f_{mel} = 2600 * \log_{10} \left(1 + \frac{z}{600} \right) \quad (7)$$

z refers to the audio frequency value based on the frequency scale, and f_{mel} refers to the audio frequency value under the Mel scale.

B. Audio Feature Selection and Evaluation

Different audio signal processing methods would obtain different features. The proposed feature extraction and classification model provides higher accuracy in music classification [19]. The early development of musical features underwent three stages of classification. The first time it was divided into four parts: semantic features, short-term features, component features, and long-term features. The second time is a more detailed division of music features based on the first time, which is divided into tone, timbre, and rhythm. The third time is based on the ideas of physics, proposing two concepts: acoustic features and perceptual features. The concepts of several classified music features mentioned above provide a direction for future research. There are now many methods to classify music features in a more detailed and clear manner, which can better reflect the similarities and differences between genres. The most common features currently are time-domain features, cepstrum features that are close to human auditory perception, and frequency domain features.

As the preferred choice for audio feature processing, time-domain features require less computation and the extraction steps are not complex. Common time-domain characteristics include short-time Zero-crossing rate (ZCR), and the calculation formula is:

$$ZCR = \frac{1}{2} \sum_{i=1}^{I-1} |\text{sgn}[y(m+1)] - \text{sgn}[y(m)]| \quad (8)$$

Among them, $y(m)$ refers to the discrete signal, and the $\text{sgn}(y)$ uncton is as:

$$\text{sgn}(y) = \begin{cases} 1 & y > 0 \\ 0 & y = 0 \\ -1 & y < 0 \end{cases} \quad (9)$$

e frequency domain feature acquisition step first utilizes Fourier transform to transform music signals from time domain to frequency domain, and then performs statistical analysis and calculation on the signals within the frequency domain. The commonly used parameters for frequency domain features include SE (Spectral Entropy), SF (Spectral Flux), Spectral Centroid (SC), and Spectral Rolloff (SR).

The spectral entropy SE represents the relationship between power spectrum and entropy rate, and the formula is:

$$SE = - \sum_{F=0}^{\frac{F_w}{2}} Q(f) \log_2 [Q(f)] \quad (10)$$

Among them, $Q(f)$ refers to the power spectral density and F_w refers to the sampling frequency.

The spectrum represents the frequency distribution of an audio signal, and the spectral centroid is a way to represent the center of the spectrum. The number of high-frequency components of an audio signal is directly proportional to the value of the spectral centroid. The formula for the spectral centroid is:

$$SC = \frac{\sum_{w=i}^h w |F(w)|^2}{\sum_{w=i}^h |F(w)|^2} \quad (11)$$

The formula for spectral flux SF is:

$$SF = \frac{1}{h-i} \sum_{w=i}^h |F(w+1) - F(w)| \quad (12)$$

V. EXPERIMENT ON CLASSIFYING ELECTRONIC MUSIC GENRES

In order to verify the genre classification performance of each model, a dataset publicly available on W website was selected, and 2700 different types of electronic music were selected as the experimental data for this article. Among them, there were 300 tracks each for Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Electro, and Glitch. Each training sample was monophonic, 100s in duration, and sampled at 20,050 Hz. After the experimental data was prepared, the samples were input into the CNN model, the PMG-Net electronic music genre classification model and the traditional classification model for training. The optimizer selected Adam (Adam Optimizer), and the loss function selected Cross Entry. The learning rate was set to 0.0003; the training round was set to 12, and the batch size was set to 128. Finally, the graphics processing unit (GPU) was responsible for accelerating the training. The traditional classification model, CNN model, and PMG-Net electronic music genre classification model were set as T1, T2, and T3, respectively. The classification performance of each model is shown in Fig. 4.

Fig. 4 shows the classification performance results of the three models after 12 training sessions, which is equivalent to the corresponding model accuracy. The higher the accuracy, the better the model performs in classifying electronic music genres. Among them, the accuracy distribution of the T1 model was between 0.3 and 0.53, and the best classification performance occurred during the 6th training session; the accuracy distribution of T2's model was between 0.3 and 0.62, and the best classification performance occurred during the first training; the accuracy distribution of the T3 model was between 0.542 and 0.75, and the best classification performance occurred during the 8th training. In addition, in 12 model training sessions, if the accuracy is greater than or equal to 0.5, T1 would be trained 2 times; T2 would be trained 4 times, and T3 would be trained 12 times. The training results indicated that among these three models, T3 had the best classification performance and the highest accuracy for electronic music genres.

Fig. 5 reflects the classification error changes of the three models with different training rounds. During the training process, the errors in the previous rounds of training were generally relatively large. T1 and T2 approached convergence after six rounds of training, and the classification error gradually decreased to less than 1. T3 approached convergence after four rounds of training, and the classification error gradually decreased to less than 1. Moreover, the error of each round of T3 training was smaller than T1 and T2. The fluctuation range of T3 error was relatively small, and the classification was relatively stable.

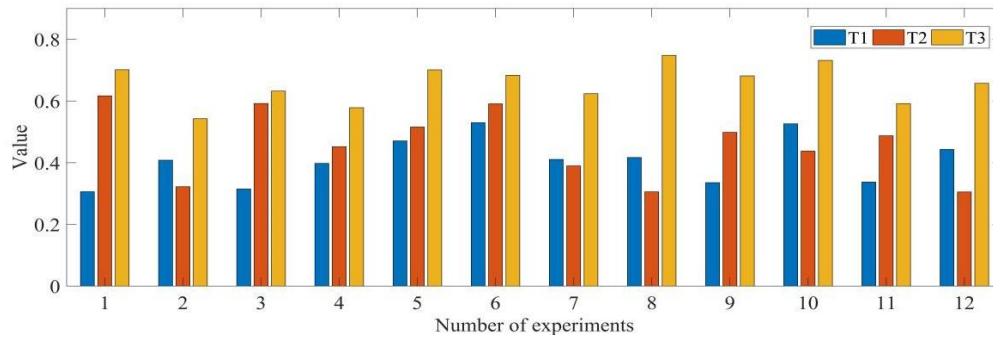


Fig. 4. Classification performance of three models trained 12 times.

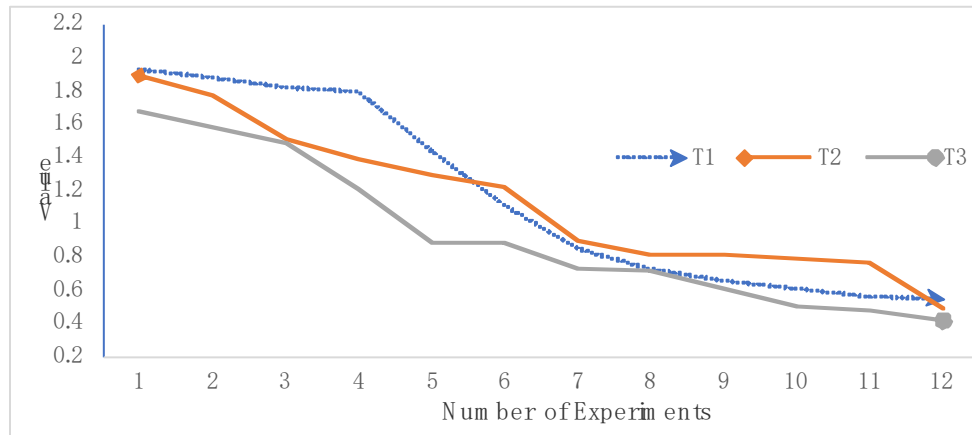


Fig. 5. Changes in classification error with training rounds.

To classify electronic music genres, it is not only necessary to consider classification accuracy, but also to evaluate the model’s signal processing speed and feature recognition speed for audio samples. Signal processing was conducted using 2700 pre experimental mono audio samples. It was known that each sample had a duration of 100s and a sampling frequency of 20050 Hz. 12 batches of training were conducted, and the average speed of the total number of times was taken as the processing result, as shown in Table I.

In Table I, T1’s audio signal processing duration was 53.92s-59.94s, and the fastest electronic music type was Techno; T2’s audio signal processing duration was 45.99s-54.88s, and the fastest electronic music type is the Trip-Hop; the audio signal processing duration of T3 was 30.22s-39.02s,

and the fastest electronic music type was Glitch. The experimental results showed that T1 and T2 were not as fast as T3 in audio signal processing, and T3 had good processing performance.

Audio signals have short-term stationarity. Generally, the processing and analysis of audio signals are based on “short-term”, which indicates that their characteristics remain unchanged within a short time range (10s-30s). Therefore, when extracting the features of audio signals, the signal is segmented and windowed, and the results at each frame are actually the return value of the feature extraction function. Using the experimental data and conditions of the audio signal mentioned above to collect feature extraction speed, the results are shown in Fig. 6.

TABLE I. SIGNAL PROCESSING SPEED OF THREE MODELS FOR AUDIO SAMPLES (IN SECONDS)

	Dubstep	Trip-Hop	Ambient	House	Techno	Trance	Disco	Electro	Glitch
T1	59.83	54.33	59.74	58.3	53.92	59.94	57.49	55.63	55.89
T2	46.81	45.99	53.46	53.65	48.24	47.65	47.18	54.04	54.88
T3	30.73	32.57	32.79	34.67	30.69	31.96	37.62	39.02	30.22

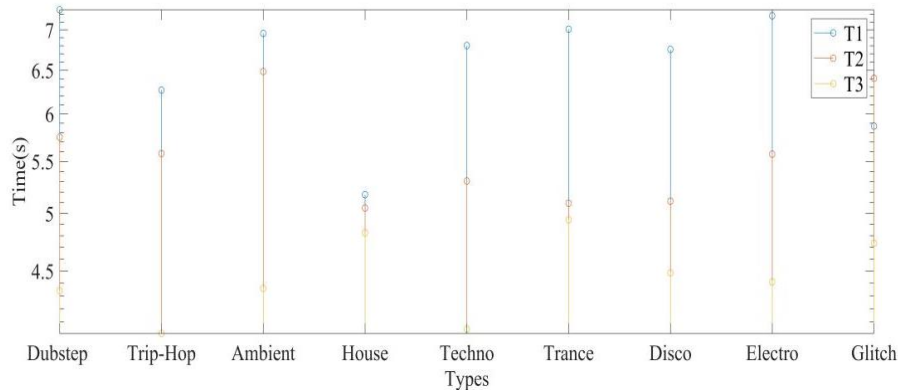


Fig. 6. Audio feature extraction speed under three models.

In Fig. 6, for nine types of audio samples, including Dubstep, Trip-Hop, Ambient, House, Techno, Trance, Disco, Electro, and Glitch, the feature extraction speed of T1 was the fastest at 5.18 seconds, with most feature extraction speeds maintained between 6-7 seconds; the feature extraction speed of T2 for these nine types of audio samples was the fastest at 5.05 seconds, with most feature extraction speeds maintained between 5-6 seconds; the fastest feature extraction speed for these nine types of audio samples in T3 was 4.01 seconds, and the overall feature extraction speed remained between 4-5 seconds. The feature extraction speed of T3 was faster than that of T1 and T2.

The traditional electronic music genre classification model had unstable performance, low classification accuracy, and single music signal features. The CNN model took the second place, and the PMG-Net electronic music genre classification model had the best classification effect on electronic music genres.

In order to further illustrate the classification effect of the PMG-Net electronic music genre classification model chosen in this paper, firstly, without any comparable results, 100 electronic music samples of different genres were randomly selected from the publicly available dataset of W network for 10 experiments. Among them, 20 samples each of reverberating heavy beat Dubstep, godly dance Trip-Hop, ambient Ambient, hoedown House, and techno were selected. The results are shown in Table II.

According to the results in Table II, it is found that the PMG-Net electronic music genre classification model classifies the accurate number very close to the standard classification number. The high degree of accuracy close to the standard number of classifications means that the model is more capable of generalization. Even when confronted with new, unseen music samples, the model is able to accurately categorize them into the correct genre. This suggests that the model does not just memorize the training data, but learns the universal music genre characteristics.

Secondly PMG-Net electronic music genre classification model in the case of comparing with convolutional neural network model and traditional classification model, the classification results of convolutional neural network model and traditional classification model are shown in Table III and Table IV.

TABLE II. PMG-NET ELECTRONIC MUSIC GENRE CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	20	20	20	19	19
2	20	20	20	20	20
3	20	20	20	20	20
4	20	19	20	20	19
5	19	20	20	20	20
6	20	19	19	20	19
7	20	20	20	20	20
8	19	20	20	20	20
9	20	20	20	19	20
10	20	20	19	20	19

TABLE III. TRADITIONAL CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	11	11	16	15	14
2	12	10	12	14	16
3	11	11	17	11	16
4	11	14	13	16	12
5	11	17	16	15	10
6	10	11	16	12	15
7	11	14	12	12	10
8	16	14	11	15	17
9	16	12	13	16	16
10	17	11	11	10	16

According to the results in Table III, it was found that the number of correct classifications for the traditional classification model ranged from 10 to 17. There are errors in every music classification.

TABLE IV. CNN CLASSIFICATION MODEL NUMBER OF CORRECT CLASSIFICATIONS (IN SAMPLES)

	Dubstep	Trip-Hop	Ambient	House	Techno
1	11	18	16	15	14
2	12	16	12	14	16
3	18	11	17	18	16
4	11	14	18	16	12
5	18	17	16	15	17
6	13	18	16	12	15
7	11	14	12	12	17
8	16	14	11	15	17
9	16	18	13	16	16
10	18	11	11	11	16

According to the results in Table IV, it was found that the number of correct classifications of the convolutional neural network model ranged from 11 to 18. It can be seen that the PMG-Net electronic music genre classification model has particularly obvious classification advantages both in comparison with other existing classification models and in direct experiments.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, a classification model was constructed based on deep neural network, and the publicly available dataset of W network was selected as the experimental sample. The sample data were input into the convolutional neural network model, the PMG-Net electronic music genre classification model, and the traditional classification model for training, respectively. The experiments were conducted in the following four aspects: in the aspect of the model classification performance, the classification performance of PMG-Net electronic music genre classification model was the best, and the training accuracy was high in every round. In terms of model classification performance, the PMG-Net electronic music genre classification model had the best classification performance, with a high accuracy of more than 0.5 in each round of training; in terms of model classification error, the PMG-Net electronic music genre classification model had a small classification error and was stable; in terms of audio signal processing, the PMG-Net electronic music genre classification model had the fastest processing speed, with a processing time of less than 40 seconds for each round of processing; in terms of feature extraction for audio samples, both the convolutional neural network model and the traditional classification model had the fastest feature extraction speed. Classification models had feature extraction speeds of more than five seconds, and only the PMG-Net electronic music genre classification model had a feature extraction speed of less than five seconds.

In conclusion, using deep neural networks to customize the PMG-Net model for electronic music genre classification provides higher classification accuracy, automated processing, flexibility, scalability, real-time performance, and personalized classification results. These advantages make the

PMG-Net model a powerful tool for handling electronic music genre classification tasks.

However again, because there may be some fuzzy boundaries between electronic music genres, that is, certain music may be characterized by more than one genre at the same time, it is difficult to categorize them unambiguously. This requires the PMG-Net model to have some flexibility and robustness to handle diversity and ambiguity.

VII. CONCLUSION

Electronic music is particularly popular in the world today due to its vibrant rhythm and diverse forms of expression, and is increasingly favored by the public. Therefore, it is very necessary to classify electronic music genres, which can achieve comprehensive retrieval and meet the music needs of different people. People can experience the ultimate charm brought by electronic music. The PMG-Net electronic music genre classification model customized by using deep neural network had better results than the convolutional neural network model and traditional classification model in the experiments of model classification performance, classification error, audio signal processing and feature extraction speed. It can be seen that the PMG-Net electronic music genre classification model customized by deep neural network is very suitable for classifying electronic music genres, fast and accurate.

The research approach of using DNN for electronic music genre classification has certain reference value for future research on automatic classification of other music genres. The drawback of this article is that the experimental electronic music audio sample size is relatively small. Therefore, future work needs to continue to collect more available electronic music samples and input them into the model to improve the model's generalization ability. The architecture and parameter settings of the deep neural network-based PMG-Net classification model have been explored and improved to improve classification performance. Different network architectures, activation functions, loss functions, etc., or techniques such as transfer learning and integrated learning can be tried to further improve the model performance.

ACKNOWLEDGMENT

2020 science and technology research fund project of Jiangxi Provincial Department of education "Research on LBG electronic music detection method based on computer-aided technology" (Project No: GJJ209912).

REFERENCES

- [1] Zhang, K. Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network. *Wireless Communications and Mobile Computing*, 2021, 2021(4): 1-7.
- [2] Pelchat, N., Gelowitz, C. M. Neural network music genre classification. *Canadian Journal of Electrical and Computer Engineering*, 2020, 43(3): 170-173.
- [3] Dong, A. M., Liu, Z. Y., Yu, J. G., Han, Y. B., Zhou, Q. Automatic classification of music genres based on visual transformation networks. *Computer Applications*, 2022, 42(S01): 54-58.
- [4] Oramas, S., Barbieri, F., Nieto, Caballero O., Serra, X. Multimodal deep learning for music genre classification. *Transactions of the International Society for Music Information Retrieval*. 2018, 1 (1): 4-21.

- [5] Elbir, A., Aydin, N. Music genre classification and music recommendation by using deep learning. *Electronics Letters*, 2020, 56(12): 627-629.
- [6] Liu, W., Wang, J., Qu, H., Dong, L., Cao, X. Music genre classification algorithm based on spectral space domain feature attention. *Computer Applications*, 2022, 42(7):2072-2077.
- [7] Fan, S. H. Music genre classification based on improved BP neural network. *Software Engineering*, 2021, 24(9): 17-20.
- [8] Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., et al. Shortcut learning in deep neural networks. *Nature Machine Intelligence*, 2020, 2(11): 665-673.
- [9] Liu, C., Arnon, T., Lazarus, C., Strong, C., Clark, B., Kochenderfer, M. J. Algorithms for verifying deep neural networks. *Foundations and Trends® in Optimization*, 2021, 4(3-4): 244-404.
- [10] Durstewitz, D., Koppe, G., Meyer-Lindenberg, A. Deep neural networks in psychiatry. *Molecular psychiatry*, 2019, 24(11): 1583-1598.
- [11] Bau, D., Zhu, J. Y., Strobelt, H., Lapedriza, A., Zhou, B., Torralba, A. Understanding the role of individual units in a deep neural network. *Proceedings of the National Academy of Sciences*, 2020, 117(48): 30071-30078.
- [12] Li, S., Jia, K., Wen, Y., Liu, T., Tao, D. Orthogonal deep neural networks. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 43(4): 1352-1368.
- [13] Wen, Z. Y. Research on music style classification model based on swarm intelligence optimization neural network. *Modern electronics*, 2019, 42(21): 82-85.
- [14] Elbir, A., Aydin, N. Music Genre Classification and Music Recommendation by Using Deep Learning. *Electronics Letters*, 2020, 56(12): 627-629.
- [15] Du, W., Lin, X., Sun, J. W., Yu, B., Yao, K. F. An automatic music classification method based on hierarchical structure. *Miniature Microcomputer Systems*, 2018, 39(5): 888-892.
- [16] Nam, J., Choi, K., Lee, J., Chou, Y. S., Yang, H. Y. Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach. *IEEE signal processing magazine*, 2018, 36(1): 41-51.
- [17] Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., Muller, K. R. Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 2021, 109(3): 247-278.
- [18] Su, J., Vargas, D. V., Sakurai, K. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 2019, 23(5): 828-841.
- [19] Thiruvén, G. R. Speech/music classification using PLP and SVM. *International Journal of Engineering and Computer Science*, 2019, 8(2): 24469-24472.

Automatic Layout Algorithm for Graphic Language in Visual Communication Design

Xiaofang Liao¹, Xinqian Hu²

Intelligent Information Research Institute, South China Business College,
Guangdong University of Foreign Studies, Guangzhou 510545, China¹
Hunan University of Technology, Zhuzhou 412000, China²

Abstract—As computer technology advances, people's capacity for visual perception grows better, and the demands placed on computerized layouts progressively rise. The simple style of graphics is no longer the only option for computer figure video creation; instead, there is a greater tendency to visually represent the effect and improve the aesthetics and expressiveness of visuals and images. Graphic language uses visual components, including shapes, colors, typographies, images, and icons, in a visual communication context to express messages, ideas, and emotions. Graphics language encounters greater chances and obstacles against the backdrop of this information era. Consequently, it is crucial to convert data into graphics language. Visual communication is evolving in some potential directions with the advancement of technological advances and cultural convergence. The graphic language has its distinct visual meaning, and each person's visual experience is extremely diverse and exists in life with various visual elements in different layouts. A hybridized Grid and Content-based Automatic Layout (HGC-AL) algorithm for graphic language in Visual Communication Design (VCD) has been developed to produce visually balanced layouts and establish a structured system for arranging content elements. The content-based layout uses design constraints for better alignment and avoids conflict loss. The hierarchical arrangement of graphic elements in a grid layout analyzes the types of visual elements like image, text, and color. Finally, graphic language enhances the visual score and gives flexibility by allowing changes and modifications within the grid layout. Following the design requirements change, the responsive fluid grid supports various graphical content, sizes, and alignments. Thus, compared with existing layout algorithms, the proposed algorithm is validated with metrics like Intersection of Union (IoU), alignment accuracy, content coverage ratio, visual score, scalability ratio, and overall layout quality.

Keywords—Graphic language; visual communication design; layout algorithm; design elements; grid layout; content layout

I. INTRODUCTION

Generally, graphical languages [1] outperform plain languages in features and embody brevity, openness, simplicity, and ease of understanding. That is unaffected by geographical or cultural disparities and allows for vast discussions and inventions across the platform. Visual language [2] is crucial for comprehending the natural environment and human culture, particularly in an era with many visual elements. Visual examination of relational data is essential in most practical analytics applications. The automatic arrangement is a crucial prerequisite for such information to be displayed visually well. Visual communication [3] design is a

proactive behavior in the graphic form used to spread specific ideas. Most portions rely on perception and are symbolized by two-dimensional pictures, such as those seen in electronic devices, typesetting, artwork, logos, and graphic layouts [4].

The ability of visual technologies for communication to measure information quickly and in real-time has led to its widespread application in many fields in recent years [5]. Related to this, Zhao [6] investigated and examined the visual alignment of the graphic language in the layout of animated visual guiding technology and incorporated alongside orientation within the film's animation design to make the animation design more in line with recent times. Designers need to be aware of cultural considerations and ensure that graphics are understandable to various people. Liu [7] utilized various cutting-edge science and technology techniques, such as VCD, which is necessary when creating interactive animation special effects to increase the view of VCD success while achieving outstanding knowledge dissemination impacts. This article initially introduced the technology for visual communication design and special animation design. In research [8], color, space forming, and graphical-text conformance, Organizations must strengthen commercial exhibits while laying out, innovating, and reconstructing these three stages of brand design. In visual design, striking the correct balance between computerized analysis and human input is critical. Lu and Huang [9] tried to improve the visual communication of artificial intelligence in the recognition and assessment of graphic design language. With the help of significant targeted location recognition and sentiment categorization using a weighted loss convolutional neural network, it focused on enhancing segmentation of image accuracy and assessing emotions depicted in graphics design. The qualities of a graphic language are: intuitive, vivid, symbolic, and aesthetic. A single graphic language, however, has not produced superior outcomes. Hence, Tao [10] examined the impact of graphic language auto layout on the design of visual communication and comprehended the method of graphic language auto layout to support the development of visual communication. Designers must optimize the layout because the intro is essential in addition to the logo and the best placement of associated components, such as the subject matter. With the help of contrast, proportion, white space, and other elements, the formally established aesthetic rule of layout reflects an abstract generalization of the aesthetic rule of design layout. Due to the existing constraints of artificial intelligence technology, a matching framework based on the formal aesthetic law of layout is a more sophisticated abstracted

matching paradigm that necessitates parametric by the aesthetically pleasing law [11]. Visualizing relational information in visual applications provided a fast compound graph with the ability to support user-specified positioning limitations and different kinds of classifications based on previous compound spring embedding technology and best for interactive applications by combining quality layouts and spectral graph velocity. It is crucial to consider the algorithm's capability to manage various user requirements and guarantee correct enforcement [12]. In that point, to deliver various information to users, the long-known symbols supply the fundamental requirements for cutting-edge, interdisciplinary research. In VCD, a particular graphic pattern is frequently made up of just one component, and depending on how it is assembled and used, it can effectively communicate various messages to various people [13]. The messages may involve graphics and pictures used while creating advertisements, poster designs, websites, and logos, particularly in commercial photography. Designers can successfully include product features using hand-drawn techniques, graphics, and picture software, increasing people's appeal [14]. Cui et al. [15] introduced a framework for automatically producing infographics, a subset of graphic language, providing evidence suggesting that it converts a natural language expression on a piece of appropriate information into a collection of expert infographics with different layouts, colors, designs, and options that ordinary people may choose or modify according to their individual choices. However, it can only deal with a limited collection of information, bad icon matching with small font sizes, and arises ambiguity problems. Graphic language plays a relatively minor part in information transmission due to its relative lack of precision, making it difficult to pinpoint information precisely. To avoid this, Wu and Fang [16] suggested a visual communication system based on the hyperspectral processing of images using entropy control stability with good visual communication effects. They can successfully increase the visual communication design effect. Existing methods have looked into using a force simulation and proxy geometry to simplify handling collisions for irregular forms. These specialized force-directed layouts are frequently unstable and necessitate additional constraints to function correctly. Kristiansen et al. [17] described a method for locating central components in a unique grid layout to provide equivalent attractive forces to address these deficiencies. Dayama et al. [18] proposed an integer programming method for interactive layout transfer, where the layout of a source design is automatically transferred using a chosen reference pattern layout while adhering to applicable rules. The source design is an initial rough working draught converted into a rough draught into the final objective layout using a reference pattern. The major objective of the study is:

- The aesthetics and visual attractiveness score of graphic language layouts are enhanced by using the HGC-AL algorithm.
- Apply design constraints for evaluating alignment accuracy, prevent conflict, and effectively design the layout using visual graphics elements.
- Provide a visual balance score by maintaining the hierarchy of elements with their relative importance

and update the design changes using responsive fluid grid supports.

- Evaluate the proposed approach using performance metrics like IoU, content coverage ratio, alignment accuracy, visual score, scalability ratio, and overall layout quality.

The remainder of this investigation is organized as follows: Section II details the prior work on visual and graphics language-related layout implementations. Section III details the data source description and proposes the implementation of HGC-AL using various graphic languages and visual elements for effective communication design. Section IV presents the comparative analysis. Section V discusses the results. Section VI concludes the work with limitations of the current study and future work.

II. PRIOR WORK

Zhou et al. [19] proposed a Composition-aware Graphic Layout integrated with a Deep Generative Adversarial Network (CaGL-DGAN) model with an Adam optimizer used to create layouts based on the overall and spatial visual information of input images. A novel domain alignment module was created to close the mismatch between the test inputs without using masks and the inputs used for training with the hint masks. The evaluated metrics are user satisfaction, visual balance, overlapping and graphical readability with root mean square values of x and y dimensions of the text elements using the poster layout design dataset with visual text variation of 0.026 difference with a lack of focus in color base.

Yu [20] elaborated a parallel selection logo to explain the parallel procedure and selection procedure for graphic languages in VCD and realizes the automatic layout of graphical languages using the parallel selection process. Syntax description based on rules and semantic description based on abstract state machines describes visual languages. The ant colony algorithm gets the graphics' demonstrated spot in VCD. The fixed value approach utilized for figuring out the showcase size of the buffer image takes the least layout land optimum surface used as the goal. Results demonstrated that the favorable scheduling impact increased focus and satisfaction.

Wang [21] suggested an algorithm for automatically arranging graphic language in VCD with the Ant Colony (AC) algorithm to get the most effective display location of the images in VCD. The Fixed Value Method (FVM) for estimating the display dimensions of the buffer graphic takes a minimal amount of layout and the maximum surface utilization as the function's objective. Graphic language can be described using syntactic analysis according to rules and semantics. Its application in real-time or resource constraints may consume greater time for processing and demand for resources.

Ma [22] developed a graphics language-based automatic scheduling technique using the Allocation of Resources State (ARS) in automated planning research. Both service management based on port resource conservation of an optical network centered on a multimode and an optical panel allocation of resources is proposed. Focussed on the application and described the dynamic layout's design concepts, methodologies, and shapes.

Wang and Li [23] suggested an automatic structure of graphics language in VCD that successfully resolves the conflict between textural capacity for storage and realism and accomplishes real-time deconstruction and visualization using programmable graphical processors. The technique automatically inserts every element of the code database throughout the compression and coding of the image. The results showed higher satisfaction; it must be tested and validated using various design scenarios not discussed in this study to ensure the algorithm's dependability and adaptability.

Sun and Mattek [24] researched Network VCD is familiar with the design thinking processes for visual information communication and has a wide range of expressive and communicative options. The algorithm of graphic language automatic arrangement in Nvcd is also researched to determine the ideal size and position in NVCD for graphic arrangement. The experimental results showed that the design of graphics makes up 27% of the total, next to film and television design (21%), advertising design (19%), typeface design (14%), and logo design (9%). However, not every user can utilize digital platforms or comprehend nuanced visual cues.

Gong and Fang [25] built a framework for automatic layout optimization called Faster-R-CNN to identify and extract an essential element from posters. It achieves recognition and placement of the format's elements by learning and categorizing the elements using RPN. The main focus is to learn the selected basic format template with the poster case format. The skeletal labeling of the layout's elements is the primary basis for classifying those elements. This framework produced a detection rate of 95% and satisfied the visual communication needs of public cultural signs. The layout is optimized while ensuring that cultural logos adhere to the rules of visual communication and exceed the satisfaction ratio above 70% using the three-division method. It emphasizes the categorized layout's core body and numerous text sections by Soto and Yoo [26]. The exact location of each component can be determined and located once the labeling of the layout elements is complete. The approach makes it possible to learn the target case's layout and finish learning the Layout migration.

Tabata et al. [27] developed a system for automatically generating layouts that combine i) random generation, a set of minimal condition rules, and ii) an assessment of adherence to the intended design aesthetic. This technique makes it feasible to produce numerous innovative layouts while maintaining the original design aesthetic. When text or images add information to layouts, an appropriate layout that considers content and design is automatically developed. A learning-to-rank predictor evaluates the work's look, design, structure, and content; the top ratings are then given to the user. Users can significantly increase layout development and editing efficiency. It could take a lot of time and computation to generate a lot of layout candidates.

The automatic layout procedure performed by an a priori designing approach still needs human input. Thus, Huo and Wang [28] provided an Artificial Intelligence-based (AI) method for designing poster layouts. A Learner and a Generator (L-G) are both components of the layout construction

process. First, the learner generated the initial templates for various composition scenarios using the spatial transformation system. The generator optimizes the first template to create many optimum templates, relying on the LeNet architecture utilizing the golden ratio and the trilateration technique. Due to the template's structure and framing fashion, the starting points are kept in an archive of matching templates. The lack of annotated training datasets focusing on poster layouts alone might have an impact.

Zhang [29] enhanced the two-way long- and short-memory model of a neural network to identify the poster using graphic language and investigated the mathematical modeling of graphic language in VCD better, thereby increasing the accuracy of the computer's ability as 0.843% and utilized the minimal time for the recognition process. Hence, the classification efficiency here is 1692.5 s; the identified research gaps in this study are network representation ability is insufficient and the network correctness is not elevated.

Xu and Shi [30] created a collection of automatically suggested visual type annotation systems that primarily process data and look for big data in the system using user field information. The user can then be advised of the best visual mark to use so that the information can be presented to them as graphics. The attributes used for automatic recommendation are dimension, color, size, shape, label, and tag type. Users can save additional time and effort by having the system recommend the best label style in single and bi-axis types. The major limitation here is no further in-depth examination of the information. Only tag options are recommended to users.

He [31] explored the design process of a visual interface using interactivity and the experience of users to achieve the goal of information transfer by employing interactive information visualization. This article makes recommendations for improving the interactive design of visual communication from the focal points of attention, awareness, and memories by analyzing users' cognitive processes in each stage of visual-based activities. The algorithm's accuracy reaches 0.963%, and user satisfaction is 0.959%, with a faster convergence time. However, this study has drawbacks related to the detailed description of expression and visual aspects of elements.

Kikuchi et al. [32] proposed a framework name Constrained Layout Generation through Optimization (CLO) with unconstrained layout generative adversarial network++ with the latent codes for evaluating the alignment of elements, overlapping issues, and other user-specified relations. The metrics related to the Intersection of Union (IoU), constraint violation, and alignment comparing three publicly available graphic-related datasets, including various visual elements like images, text, icons, logos, etc. The major drawback is issues related to visual quality may occur there may be a need for an extra training period for discriminators.

The review study details the prior work related to the VCD field with graphic language of automatic layout algorithm with fullest satisfaction of user identification ratio and various design scenarios, practical usage, and dynamic designs.

III. PROPOSED ALGORITHM

Using graphics and visuals to communicate ideas is known as visual communication design. Four fundamental components are primarily used when producing visual communication effects: graphical style, text structure, graphic hues, and graphic arrangement layout. Providing visual information and making the viewer attractive are important aspects of layout design. Quantitative design encompasses the development of technological advances in graphics and visuals. It can separate graphic images' color and structure details into separate, quantifiable, and measurable independent pieces. Graphic language includes the principles, procedures, and phrases used in visual design to build coherent and meaningful compositions. Graphical components such as lines, forms, shades, layout, structure, and visual hierarchical arrangement express ideas and produce striking visuals. In the subject of visual communication design, messages are effectively conveyed through the use of images, typography, and layout. It can be applied in some ways, including marketing, advertising, and producing signage or educational materials. To successfully explain the facts or information being shown, visualization frequently uses design ideas from graphic languages, such as the use of a visual hierarchy, color coding, and typographical nature. The fundamental element of visual layouts, such as those for newspapers and magazines, advertisements, cartoons, and internet pages, is layout. A good layout can improve content presentation, direct reader focus, and increase visual appeal. Fig. 1 depicts the implementation of HGC-AL for effective VCD.

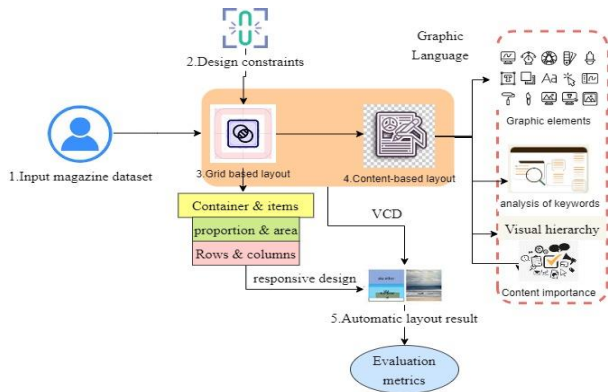


Fig. 1. Implementation of HGC-AL.

The innovative approach combines two fundamental principles - grid-based layout and content-based design - to enhance the efficiency and creativity of graphic language in visual communication. By integrating these two methods, the work offers a comprehensive solution that ensures visual coherence through the use of grids but also allows for dynamic and adaptive layouts based on the presented content. This hybridized approach empowers designers with a versatile toolset to create visually appealing and effective communication materials. Additionally, it addresses the evolving demands of contemporary design, where flexibility and automation are crucial. Overall, this work opens up new possibilities for graphic designers, enriching their creative processes and enabling them to produce more engaging and impactful visual communication designs.

A. Input Magazine Dataset

The magazine dataset includes several categories of graphic elements, including images, text, colors, and other visual components represented in this block as the input information used by the layout algorithm. However, some datasets with semantic segmentation are openly accessible to magazines and academic journals. From Zheng et al. [33], 3 919 pages of magazines from the Internet that cover six common topics, namely fashion, nutrition, headlines, technology, traveling, and weddings, make up the dataset. The six categories total 685, 753, 618, 509, 721, and 633 pages, respectively. The types of graphic language labels to encode the layout of each page to enhance VCD with element categories are Text, Image, Headlines, color, and Background. Pages from various categories display a variety of layouts in terms of the quantity, size, and spatial positioning of page elements. The most commonly used AR of magazines are 4/3; hence the general representation of the layout to be generated may be fixed as the dimension of 60×45. Capture the design aspects: Identify the visual aspects that should be properly integrated into the layout, and graphics could consist of text blocks, pictures, logos, icons, and other graphical elements. Each element should have the appropriate qualities, such as dimension, position, alignment, and visual order. Based on this dataset, the layout of the graphic visual elements needs to be arranged automatically by following the design constraints given in subsequent content and applying HGC-AL.

B. Design Constraints for Various Attributes

Before applying the layout algorithm, the basic design constraints must be fixed to enhance the visual appeal, usefulness, and user experience of the graphical layout. The goal of adding design qualities into a layout algorithm utilizing graphic language is applied in the collected dataset to generate a layout design. The design constraints are used to direct the layout-generating process represented by this block with criteria for alignment, overlapping, size, location and spacing, proportions, and other design principles. Design the guidelines that must be followed throughout the layout-generating process, including element associations or alignment, closeness, and spacing constraints. These design constraints are used for element coordination, conflict avoidance, or user-specified interaction. To identify the conflict loss between elements pairs in the layout composition.

$$Cl_{loss} = \sum_{i=1}^Z \sum_{j=i}^Z \frac{e_i \cap e_j}{e_i} \quad (1)$$

As mentioned in Eq. (1), $e_i \cap e_j$ represents the intersection or overlapping region of two elements i and j . Element alignment is adopted for the sample rectangle shape with the two basic alignment principles, edge alignment and center alignment, to avoid this conflict loss problem. A common strategy is edge alignment, which involves aligning parts with an outer boundary that intuitively matches their exterior edges. The function provides the orientation and positional information for the list elements to create acceptable visual groups. The two elements' size and shape are considered when choosing the alignment. A visual hierarchy throughout the layout structure is established using alignment. You can clearly distinguish between different subsections or stages in

importance by aligning text elements. Aligned text enhances the design's aesthetics and attractiveness with hierarchy and increases the visual score. An automatic layout plan is more aesthetically pleasing when balanced alignment conveys an emotion of order and balance.

Pseudocode-Rectangular type element alignment for enhancing the visual score

```

Input: rectangular shape as rect (length as  $l$ , height as  $h$ , width as  $w$ ),
top, left.
Output: element alignment as  $el$ 
if  $rect.l < 3$  then
  Compare  $sim(w, h)$ 
  if  $(\max(w) > \max(h))$ , then
     $el \leftarrow el_{left}$ 
    update longer  $w$  on top
  else
     $el \leftarrow el_{top}$ 
    update longer  $h$  on left
end if
else
   $rect \leftarrow bgRect(rect)$ 
  if  $rect.h > rect.w$  then
     $el \leftarrow el_{left}$ 
  else
     $el \leftarrow el_{top}$ 
  end if
end if

```

1) *Intersection over Union (IoU)*: IoU can be used to evaluate how well the projected arrangement and the actual layout agree or line up. Locate the area where the expected layout and the actual arrangement overlap, which can be achieved by locating the areas or components shared by the two layouts. Subtract the union area from the intersection area, and the calculated value indicates the degree of agreement or overlap between the expected and actual layouts.

IoU value ranges from 0 to 1, where 0 means no overlap and 1 means perfection in the overlap. For the intersection area of a design element, the notation $e_i \cap e_j$ and $e_i \cup e_j$ the intersection and union of measures can be calculated using Eqs. (2) and (3).

$$IoU_{VCD} = \frac{e_i \cap e_j}{e_i \cup e_j} \quad (2)$$

$$e_i \cap e_j = (e_{i1}^l - e_{i0}^l) * (e_{j1}^l - e_{j0}^l) \quad (3)$$

Combining the two boxes' areas and deducting them at their junction allows one to find the union area of two boxes using Eq. (4) and Eq. (5).

$$e_i = (e_{i1}^i - e_{i0}^i) * (e_{j1}^i - e_{j0}^i) \quad (4)$$

$$e_j = (e_{j1}^j - e_{j0}^j) * (e_{i1}^j - e_{i0}^j) \quad (5)$$

In visual communication design, an image's aspect ratio is useful since it offers important details about the proportions and shape of the image in three variants: portrait, landscape, and square. Suppose an image element's aspect ratio is greater than the picture's. In that case, the image is scaled to match the

image element's width while maintaining its aspect ratio, and then the centers of the two are aligned. Then align the images along the vertical axis by the center, top, or bottom boundary because the scaled image is taller than the image element using Eq. (6).

$$AR = w/h \quad (6)$$

w indicates the width of an image, h represents a height of an image.

The design attributes are included with graphic language and consider i) color as the initial graphic language representation to generate an automatic grid layout based on the input contents. The color scheme is carefully chosen to produce a unified and aesthetically pleasing appearance. Depending on the branding and content of the magazine, it might use both vivid and subdued colors. The appropriate application of color may generate feelings, convey meaning, and have a strong visual impression. The mood, accessibility, and overall aesthetic elegance of a design can all be impacted by color choices. Readability can be improved by using lighter backgrounds with darker text or the opposite. A focal point or important information can be highlighted with a strong, brilliant color to grab and direct the reader's attention. The layout automatically generates the results according to the given visual element using the "adjustColorPalatte ()" method in design constraints of the color attribute

2) *Size and location constraints*: Size and location constraints are derived from Kikuchi et al. [32] to generate an effective layout. Three categories must be considered for the size element constraint: small, large, and equal. If a user specifies that j^{th} element has to be larger than the i^{th} element, the sum of cost functions of larger relation is calculated using Eq. (7) as:

$$S_{lg}(e_i, e_j) = \max((1 + \gamma)A(e_i) - A(e_j), 0) \quad (7)$$

Where $A(\cdot)$ represents the function that the occupied area of the given bounding box and γ is a tolerating variable shared between the relations of the size. Likewise, Eq. (8) calculates the location constraint with five variables above, left, right, bottom, and overlap. If an element e_i has to be present right of an element e_j then, the right location's cost-related function is given.

$$L_r(e_i, e_j) = \max(y_l(e_j) - y_r(e_i), 0) \quad (8)$$

Where $y_l(\cdot)$ and $y_r(\cdot)$ denotes a function that returns a considered bounding box's left and right coordinates.

Based on these design constraints for aligning text, image, size, location, and color arrangements, IoU analysis inside a layout with correct columns and width spans in the container must follow the grid layout design.

C. Grid-based Layout

This block represents the grid-based layout algorithm, creating a grid structure with well-defined columns and rows out of the available space. The grid provides a structural framework for organizing and positioning the graphic languages within the layout. Grid-based layouts prioritize

structure, responsiveness, and resource optimization, while content-based layouts emphasize every component's importance and enhance the overall user experience. A grid-based layout's main goal is to arrange items methodically and consistently. It strives to align pieces along horizontal and vertical principles to create a visually balanced and well-organized composition. Grids ensure that elements keep a consistent connection to one another and give the structure an aesthetic feeling and harmony. Grid-based layouts are effective in using space, especially when dealing with a lot of content or objects arranged in a grid-like pattern. By placing components in a grid, they may maximize the potential of the space and available reduce unused space, assuring an aesthetically appealing and well-organized design. According to the screen size, the fluid grid's variable content breadth can span the entire edge concerning the graphical language used in the design constraints. Columns in a fluid grid have variable widths, whereas the gutters and side margins are fixed.

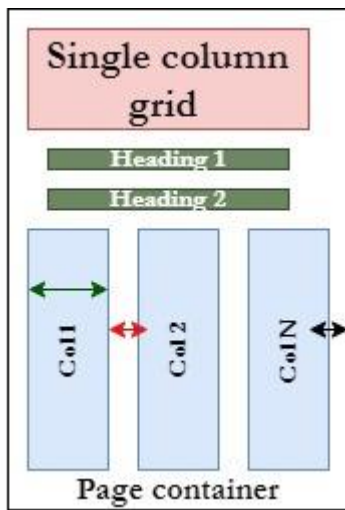


Fig. 2. Basic grid layout design for a magazine.

Create a structure based on grids with a fixed number of rows and columns, as shown in Fig. 2, with several attributes. The red color arrowhead represents the space between column grids in which the areas between the columns are known as gutters aid in dividing the text; the green arrow represents the space within the column surrounded by the page container, adopts the design layout based on the given input type of graphic language to enable the good score of visual communication. Breakpoint refers to a particular range of sizes of screens wherein the layout is re-sized to the available screen size for the optimum layout view. The margin of the layout defines the magazine content and the edge of the screen layout. The grid serves as an outline of the algorithm that gives the layout a dependable sense of alignment and structure of visual elements arranged based on a content-based layout. Here the layout dimension is taken as of width (w) size as 100 cm; hence the number of columns in a grid cell is taken as 12 to calculate the width of each column as $w/\text{no. of columns}$, then it gives a unit-wise wider area. Similarly, the no. of rows is taken as 3 for this $h/\text{no. of rows}$ is calculated; hence the resultant value is 15 units high. Hence, the resultant grid sample will be square with equal width and height space. This grid layout separates the pages into sections like rows and

columns for varying magazine trends arrangement to enhance the design principle of graphic language. Then follow the alignment rules to create visual order and balance with the help of proximities like size and location for enhancing the reader's understanding and satisfaction.

The layout design is aligned with the proper graphic language with its importance related to contents for enabling visual communication and delivering the message with good visual score and readability. The content-based layout needs to be followed, which is given in the subsequent section.

D. Content-based Layout

The content-based layout algorithm is represented by this block in which each graphic element's attributes and significance are examined, including its visual impact and relationship importance to other elements, as shown in Fig. 3. It uses this data to decide where elements should be placed in the grid and how they should be arranged. Here the importance of keywords to display the content about the given category based on term frequency-inverse document frequency is given below:

1) *Keyword/text importance*: The term frequency-inverse document frequency ($tf - idocf$) method counts the number of words in a collection of documents and calculates a score for every word to reflect its weight in the corpus and document. The term frequency calculates the word frequency in the text and is heavily dependent on the size of the content inside the document as well as generality. Since TF is specific about every document and word, the term t denotes the term or word doc denotes the document where a group of words occurred. N denotes the corpus count referring to the total document set. DF is the number of times a term appears in the document set N and identifies a document's significance within the corpus as a whole. DF is the number of times the phrase t appears in the information set N . In other words, DF represents the number of documents containing the word. IDF , which quantifies the informativeness of a term t , is the inverse of document frequency.

$$tf(t, doc) = \text{count of } t \text{ in } doc / \text{no. of words in } doc \quad (9)$$

$$idocf(t) = \log\left(\frac{N}{(docf+1)}\right) \quad (10)$$

$$tf - idocf = \left(\frac{\text{count of } t \text{ in } doc / \text{no. of words in } doc}{idocf(t)}\right) \quad (11)$$

From these above Eq. (9)-(11), the relative weightage of a word can be calculated where the repeatedly occurring words like stop-words in a magazine have a very low $idocf$ score. Depending on the high $TF-IDF$ evaluations, determine the keywords or significant terms and allocate sizes to text elements according to their significance or relevancy. For example, greater sizes can be given to those with higher $TF-IDF$ values to visually emphasize the importance of text elements within the layout. Consider utilizing bigger font sizes, bold or italics formatting, or distinctive colors to draw attention to keywords within the content. The content makes the key phrases more noticeable to readers and stimulates interaction with the destination knowledge.

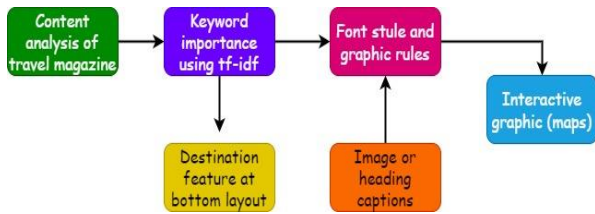


Fig. 3. Content-based layout of graphic language for traveling magazine.

E. Automatic Layout Representation

This block depicts the algorithm's final result, a created layout that blends the concepts of grid-based and content-based layouts. According to the limitations of the design and aesthetic considerations, it displays the graphic elements' ideal arrangement.

The font size for the text and text-over-image elements is fixed to make it easier to see the generated layouts. In contrast, the font size for the headline and headline-over-image portions is set to a minimum of three times larger and fluctuates depending on the size of the corresponding regions. According to the design and intended amount of emphasis, the font size commonly varies between 24 pt to 72 pt or greater. Change the font size based on the image proportions and the headlines' effect on readers. The grid structure adjusts and scales by implementing responsive fluid grids following the available screen size. The technique gives grid columns proportional widths to alter dynamically rather than set column widths. Instead of using fixed pixel values, content items within the grid are placed using relative measurements. The technique of arranging items to demonstrate their relative importance can be called visual hierarchy. Designers organize visual elements, such as menus, images, colors, texts, and symbols, so users can easily comprehend information. Designers modify users' views and direct them towards desired actions by arranging things logically and strategically. It ensures the best possible use of the available space by allowing objects to scale and reposition themselves according to the screen size.

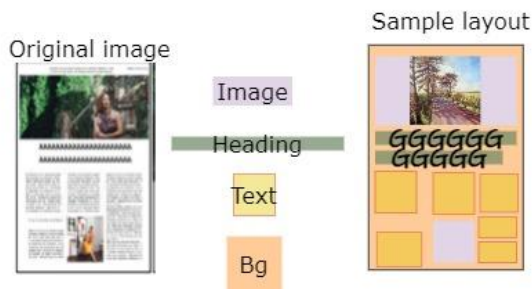


Fig. 4. Layout representation of graphic languages using visual elements like image, text, and background.

As shown in Fig. 4, the input image is taken from a magazine type called traveling, where the graphic language contains the four partitions generated to design the grid-based layout, and contents are aligned with separate columns and rows based on the width and height of the layout dimension area. Initially, the background is identified from the color palette function extracted from the color design attribute rules. The image is identified with the element pair of i , and j coordinates to find the dimension with overlap functions, size,

and location proximities. Then the heading category is aligned with boldface dark font. The text is aligned based on the left and right alignment rules without overlapping regions.

IV. EXPERIMENTAL ANALYSIS

The proposed algorithm is compared with four existing approaches, CaGL-DGAN [19], AC-FVM [21], AI-L-G [28], and CLO [32], using various metrics like IoU analysis, alignment accuracy, content coverage ratio, visual balance score, scalability ratio, and overall quality score.

A. IoU Analysis

Whenever the IoU number is high, two separate boundaries have a lot of overlap or similarities. IoU is frequently used to assess the precision of anticipated bounding boxes compared to ground truth indications in tasks like identifying visual elements and image recognition using Eq. (2).

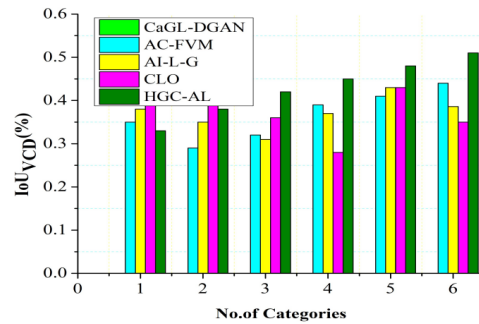


Fig. 5. Comparison analysis of IoU values.

Fig. 5 depicts the IoU analysis of image overlapping in six categories of domains like fashion, nutrition, headlines, technology, and traveling wedding. Typically, IoU is stated as a number between 0 and 1, with 0 denoting no overlap and 1 denoting a perfect match or total overlapping between the layouts. IoU can be used by designers as an input system to hone and enhance layout algorithms. Designers can pinpoint areas for improvement and modify the algorithm variables or limitations by examining the IoU ratings for several design iterations.

B. Alignment-Accuracy Measure

Using design constraints like alignment rules for graphic elements in VCD, including text or image alignment. Suppose the aligned text matches the design constraints and follows the pseudocode rules of left, right, bottom, and height categories. If the alignment is text element need to store in the fashion layout with the corresponding keywords extracted using the frequency of occurrences and its relevant scores for the proximities of size and location constraints based on the obtained score, the alignment accuracy is evaluated.

$$\text{Alignment Accuracy}(\%) = \frac{(\text{score1} * \text{wt1} + \text{score2} * \text{wt2} + \dots + \text{scoreZ} * \text{wtN})}{(\text{wt1} + \dots + \text{wtZ})} \quad (12)$$

From Eq. (12), score is calculated from relevant extraction methods like $tf - idocf$ for text, and the score will vary for other visual elements design patterns like traveling, wedding,

and technology because each magazine has different visual styles of the alignment procedure. Based on this alignment rule, the accuracy of aligned text has been calculated.

As depicted in Fig. 6, the alignment accuracy increases if the score of the extracted graphic language increases and follows the alignment rules derived from design constraints. The alignment must include the graphical language measure, like the wedding category has a separate theme like the fashion category has a separate theme in the background. Based on this graphic language analysis, the relevant text is separated and aligned based on the score acquired.

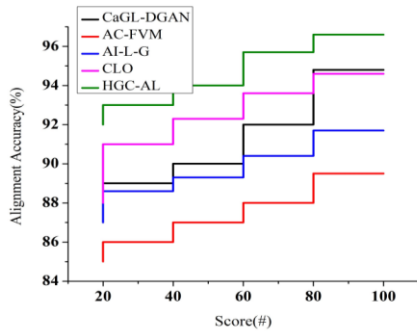


Fig. 6. Calculation of alignment accurateness.

C. Content Coverage Ratio (%)

The CCR determines the entire area that the grid's elements of content occupy, and the content coverage ratio is calculated by adding together the areas of all the content elements and dividing that sum by the grid's overall area. The CCR from Eq. (13) shows how much of the grid layout is taken up by the graphical content elements. A higher CCR indicates greater coverage of the space accessible by the design elements.

$$CCR = \frac{\text{(Sum of content element areas)}}{\text{total grid dimension}} \quad (13)$$

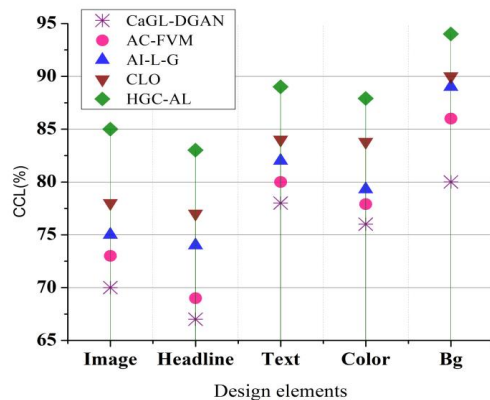


Fig. 7. CCR analysis.

As shown in Fig. 7, CCR analysis results in a comparison between the existing and proposed approach gives the better coverage aspect of design elements in HGC-AL with the additional effort of grid layout usage since this approach provides a better resource utilization related to content by identifying the relevant column and width spaces of each grid

cell to reduce the wastage of unused space. In this dataset, the magazine categories are related to several types of fashion, nutrition, and headlines, including graphic design elements like image, text, color, and background aligned with the design constraints.

D. Visual Score Analysis

As shown in Fig. 8, the graph compares the criteria aspects like proportion, text, alignment, color distribution, size, location, and image flow are examples of particular criteria that affect visual balance with various existing and proposed HGC-AL. Design best practices and these requirements should be compatible based on design constraints. For a given layout, assess each criterion separately and give it a score depending on how well it satisfies that criterion. That can be a scale of 1-10 and a subjective rating from layout designers considering the balance attained in visual communication design elements.

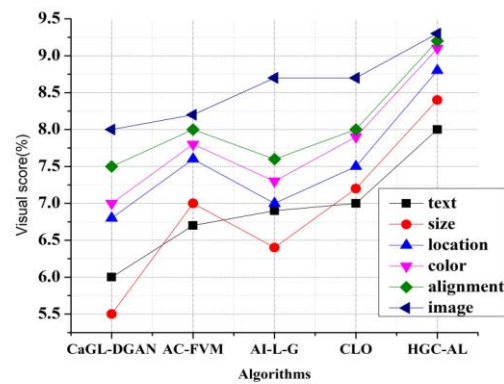


Fig. 8. Visual score comparison.

E. Analysis of Scalability Ratio (%)

An HGC-AL algorithm's scalability may also be affected by *N* no.of design constraints like conflict loss, content hierarchy, alignment rules, IoU, size, and location constraints are increasingly difficult as the amount of content rises. The algorithm's scalability can have trouble staying consistent and adhering to the rules.

From Eq. (14), the scalability of the proposed algorithm increases as the content count or size increases for six categories of magazines with different contents like image, text, color, headline, and background concerning the design constraints of the graphic language elements. Here α , β , and γ represent the coefficients that denote the relative importance of each design criterion employed in the HGC-AL algorithm.

$$\text{Scalability ratio}(\%) = \alpha * (\text{content count}) + \beta * (\text{design constraint 1}) + \dots + \gamma * (\text{design constraint N}) \quad (14)$$

From Fig. 9, the graph shows the scalability ratio analysis of various algorithms concerning design criteria or constraints and the content count of each magazine category. The proposed algorithm scales well concerning each design constraint as the content count increases. It improves the algorithm's scalability compared to other approaches using the hybridized concept.

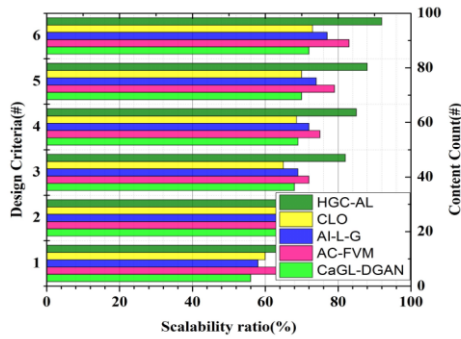


Fig. 9. Scalability ratio analysis.

F. Overall Layout Quality Consideration

The various evaluation criterions used to assess the layout quality are represented on the x-axis. Aspects like aesthetics, unity, content hierarchy, color contrast, accessibility, functionality, and other pertinent design criteria can be included in these considerations. The layout's quality score is shown on the y-axis following the evaluation criteria taken from Zheng et al. [33].

$$O_{lq} = \frac{\sum(score(i) * wt(i))}{\sum(wt(i))} \quad (15)$$

From Eq. (15), the score (i) represents the score allocated to the current i^{th} evaluation criteria, with $wt(i)$ denoting the weight assigned to the i^{th} evaluation criteria concerning the relative importance of each criterion utilized in this calculation.

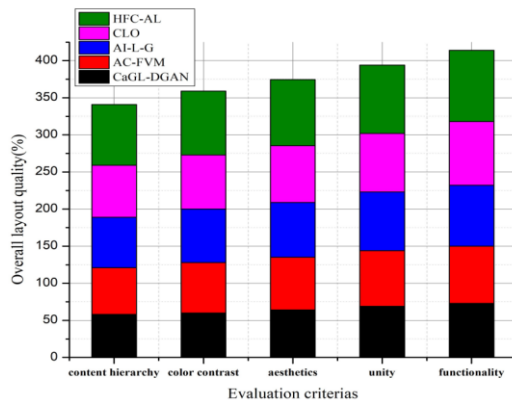


Fig. 10. Overall layout quality analysis for various algorithms.

The score calculated from Fig. 10 shows the total assessment of all design factors considering the relative importance of each parameter that produces the overall layout quality of each algorithm. By calculating this total quality score, the performance and advantages of the proposed HGC-AL technique are proven to have superior and good quality scores compared to other existing approaches.

V. DISCUSSIONS

The proposed Hybridized Grid and Content-based Automatic Layout algorithm exhibit notable advantages compared to existing methods such as CaGL-DGAN, AC-FVM, AI-L-G, and CLO in visual communication design. Its

IoU analysis scores consistently outperform these methods, indicating superior accuracy in replicating desired layouts, which is crucial for maintaining design integrity. Secondly, alignment accuracy is a standout feature, as the algorithm excels in preserving visual coherence and ensuring proper element alignment, surpassing the precision of the existing approaches. Furthermore, the content coverage ratio metric demonstrates the algorithm's proficiency in maximizing information presentation within available space, a critical aspect of effective communication design that it accomplishes more efficiently than its counterparts. The visual balance score consistently ranks higher, showcasing its aptitude for creating aesthetically pleasing designs, a feature that sets it apart in aesthetics. Its excellent scalability ratio underscores its adaptability across various design scales and complexities, making it a practical choice for various design projects. The algorithm's impressive performance across these metrics contributes to a significantly higher overall quality score than the existing methods. This comprehensive superiority positions the Hybridized Grid and Content-based Automatic Layout algorithm as a valuable and versatile tool in the field of visual communication design, offering enhanced accuracy, alignment, content coverage, aesthetics, and scalability for designers seeking to create impactful and visually engaging communication materials.

VI. CONCLUSION

The creation of images and visuals for computers not only advances computer development but also helps to improve graphic and image design. The study of VCD has changed people's conceptions of life and consumption while introducing them to innovative and personalized visual appraisal activities. For creating visually balanced layouts in Visual Communication Design (VCD), the HGC-AL algorithm offers a dependable and effective solution. The method ensures that graphic elements are placed in the best possible location while considering design limitations and alignment by combining grid-based structure with content-based analysis. It improves the visual score and offers versatility and adaptation to shifting design specifications. Metrics showing the algorithm's performance over competing layout algorithms include the intersection of union (IoU), alignment accuracy, content coverage ratio, visual score, scalability ratio, and overall layout quality. The HGC-AL algorithm improves graphic language and communication in visual design by producing aesthetically appealing and efficient layouts. Thus HGC-AL enables a more efficient design workflow by automatically ordering and placing content items in a grid layout according to established design constraints and techniques. The major limitation of this study involves that if the design constraints of graphical elements are not satisfied, it may affect the quality of the generated layout model. Hence, future work will focus on dynamic updates of the design constraints and the suitable optimization algorithms for attaining effective layout quality.

REFERENCES

[1] Z. Wang, J. Yu, A. W. Yu, Z. Dai, Y. Tsvetkov, and Y. Cao, "Simvlm: Simple visual language model pre-training with weak supervision," Openreview.net. https://openreview.net/pdf?id=GUrhfTuf_3 (accessed Aug. 24, 2023).

- [2] C. Weninger, "Multimodality in critical language textbook analysis," *Lang. Cult. Curric.*, vol. 34, no. 2, pp. 133–146, Jul. 2021, doi: 10.1080/07908318.2020.1797083.
- [3] J. E. Lee, S. Hur, and B. Watkins, "Visual communication of luxury fashion brands on social media: effects of visual complexity and brand familiarity," *J. Brand Manag.*, vol. 25, no. 5, pp. 449–462, Jan. 2018, doi: 10.1057/s41262-018-0092-6.
- [4] G. Aiello and K. Parry, *Visual Communication: Understanding Images in Media Culture*. Thousand Oaks, CA: SAGE Publications, 2020.
- [5] J. Park and B.-U. Lee, "Color image enhancement with high saturation using piecewise linear gamut mapping," *J. Vis. Commun. Image Represent.*, vol. 67, no. 102759, p. 102759, Feb. 2020, doi: 10.1016/j.jvcir.2020.102759.
- [6] L. Zhao, "The application of graphic language in animation visual guidance system under intelligent environment," *J. Intell. Syst.*, vol. 31, no. 1, pp. 1037–1054, Jan. 2022, doi: 10.1515/jisys-2022-0074.
- [7] X. Liu, "Animation special effects production method and art color research based on visual communication design," *Sci. Program.*, vol. 2022, pp. 1–13, Mar. 2022, doi: 10.1155/2022/7835917.
- [8] Z. Cao, "The application of intelligent generation technology in the visual communication design of exhibition brand," *Wirel. Commun. Mob. Comput.*, vol. 2023, pp. 1–11, Apr. 2023, doi: 10.1155/2023/1550761.
- [9] . Lu and L. Huang, "Exploration and application of graphic design language based on artificial intelligence visual communication," *Wirel. Commun. Mob. Comput.*, vol. 2022, pp. 1–10, Sep. 2022, doi: 10.1155/2022/9907303.
- [10] P. Tao, "Study on the influence of automatic layout of graphic language on visual communication design," in *Lecture Notes in Electrical Engineering*, Singapore: Springer Nature Singapore, Jan. 2022, pp. 1841–1846.
- [11] H. Yang and W. H. Hsu, "Vision-based layout detection from scientific literature using recurrent convolutional neural networks," in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021.
- [12] H. Balci and U. Dogrusoz, "FCoSE: A fast compound graph layout algorithm with constraint support," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 12, pp. 4582–4593, Dec. 2022, doi: 10.1109/tvcg.2021.3095303.
- [13] W. Yue, "Research on innovative application of new media in visual communication design," *J. Phys. Conf. Ser.*, vol. 1550, no. 3, p. 032146, May. 2020, doi: 10.1088/1742-6596/1550/3/032146.
- [14] M. Rizk, A. Baghdadi, M. Jezequel, Y. Mohanna, and Y. Atat, "No-instruction-set-computer design experience of flexible and efficient architectures for digital communication applications: two case studies on MIMO turbo detection and universal turbo demapping," *Des. Autom. Embed. Syst.*, vol. 25, no. 1, pp. 1–42, Jan. 2021, doi: 10.1007/s10617-021-09245-x.
- [15] W. Cui et al., "Text-to-viz: Automatic generation of infographics from proportion-related natural language statements," *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 1, pp. 906–916, Jan. 2020, doi: 10.1109/tvcg.2019.2934785.
- [16] D. Wu and Z. Fang, "Hyperspectral image processing technology and its application in visual communication design," *Adv. Multimed.*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/1272039.
- [17] Y. S. Kristiansen, L. Garrison, and S. Bruckner, "Content-driven layout for visualization design," in *Proceedings of the 15th International Symposium on Visual Information Communication and Interaction*, New York, NY, USA: ACM, 2022.
- [18] N. R. Dayama, S. Santala, L. Brückner, K. Todi, J. Du, and A. Oulasvirta, "Interactive Layout Transfer," in *26th International Conference on Intelligent User Interfaces*, New York, NY, USA: ACM, 2021.
- [19] M. Zhou, C. Xu, Y. Ma, T. Ge, Y. Jiang, and W. Xu, "Composition-aware graphic layout GAN for visual-textual presentation designs," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, California: International Joint Conferences on Artificial Intelligence Organization, 2022.
- [20] D. Yu, "Automatic layout algorithm of graphic language in visual communication design," in *Application of Intelligent Systems in Multimodal Information Analytics*, Cham: Springer International Publishing, 2021, pp. 930–934.
- [21] Y. Wang, "Automatic layout algorithm of graphic language in visual communication design," in *Lecture Notes in Electrical Engineering*, Singapore: Springer Nature Singapore, 2022, pp. 1529–1533.
- [22] Z. Ma, "Application of graphic language automatic arrangement algorithm in the design of visual communication," in *Lecture Notes on Data Engineering and Communications Technologies*, Singapore: Springer Singapore, 2022, pp. 60–66.
- [23] Y. Wang and X. Li, "Research on automatic layout algorithm of graphic language in visual communication design under new media context," in *2021 4th International Conference on Information Systems and Computer Aided Education*, New York, NY, USA: ACM, 2021.
- [24] H. Sun and A. Mattek, "Diversified characteristics and performance of network visual communication design based on internet technology," in *Lecture Notes in Electrical Engineering*, Singapore: Springer Nature Singapore, 2022, pp. 1165–1172.
- [25] X. Gong and J. Fang, "Design of public cultural sign based on Faster-RCNN and its application in urban visual communication," *PeerJ Comput. Sci.*, vol. 9, no. e1399, p. e1399, Jun. 2023, doi: 10.7717/peerj-cs.1399.
- [26] C. Soto and S. Yoo, "Visual detection with context for document layout analysis," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2019.
- [27] S. Tabata, H. Yoshihara, H. Maeda, and K. Yokoyama, "Automatic layout generation for graphical design magazines," in *ACM SIGGRAPH 2019 Posters*, New York, NY, USA: ACM, 2019.
- [28] H. Huo and F. Wang, "A study of artificial intelligence-based poster layout design in visual communication," *Sci. Program.*, vol. 2022, pp. 1–9, Jun. 2022, doi: 10.1155/2022/1191073.
- [29] B. Zhang, "Graphic language representation in visual communication design based on two-way long- and short-memory model," *Math. Probl. Eng.*, vol. 2022, pp. 1–8, Jun. 2022, doi: 10.1155/2022/6032255.
- [30] H. Xu and L. Shi, "Analysis of graphic language expression in visual communication design," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–7, Sep. 2022, doi: 10.1155/2022/2434992.
- [31] X. He, "Interactive mode of visual communication based on information visualization theory," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–10, Jul. 2022, doi: 10.1155/2022/4482669.
- [32] K. Kikuchi, E. Simo-Serra, M. Otani, and K. Yamaguchi, "Constrained graphic layout generation via latent optimization," in *Proceedings of the 29th ACM International Conference on Multimedia*, New York, NY, USA: ACM, 2021.
- [33] X. Zheng, X. Qiao, Y. Cao, and R. W. H. Lau, "Content-aware generative modeling of graphic design layouts," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–15, Jul. 2019, doi: 10.1145/3306346.3322971.

Smart Sensor Signal-Assisted Behavioral Model and Control of Live Interaction in Digital Media Art

Pujie Li*, Shi Bai

School of Art and Design Bengbu University
Bengbu Anhui, 233000
China

Abstract—Digital media art immersive scene design is a type of art design based on the theory of positive psychology mind flow theory, using digital media as the main technology and tool to build a certain scene by stimulating the senses and perception of the user so that they can achieve a state of immersion and forgetting other things. In this paper, we discuss the application of digital experience technology in designing art scene interaction devices by combining intelligent sensor signal analysis with multimodal interaction. Based on this, a new inductive displacement sensing element is proposed, which adopts square wave driving mode and op-amp circuit to extract signals, overcoming the shortcomings of the traditional inductive displacement sensing element, gaining the advantages of small size, lightweight, good linearity, high-frequency response, a simple driving circuit, and signal detection circuit, and more easily adaptable to microcomputer control. A more comprehensive anti-interference and system fault self-diagnosis design is carried out for the sensor system to ensure the stability and reliability of the system. An intelligent digital filtering algorithm with program judgment is proposed, with better smoothing ability and faster response speed. The multimodal interaction in digital experience design strategy is applied to the design practice, and a series of diversified device design solutions are proposed suitable for on-site interaction behavior.

Keywords—Intelligent sensors; digital media; VR technology; artistic interaction

I. INTRODUCTION

Under the development of art and culture globalization, the public's diversified needs for art and its dissemination are gradually growing rapidly. As a special position for the development of communication and art, art needs to meet the various needs of various audience groups. The value of art is mainly reflected in the artists' hard work, and the public's positive feedback on art and culture. It is imperative that the use of digital technology is no longer superficial and requires a more professional and rigorous attitude to regulate the use of new media technologies [1]. In the face of the rapid development of science and technology and digital media in the information age, cross-discipline, interdisciplinary, and cross-media have become mainstream topics in the study of art and technology and an indispensable part of the art development process. Artists and technology practitioners have become very interested in the integration of new technologies and new media [2], exploring new ideas and new problems that may arise from the union of art and technology, and making efforts to explore this. However, due to the late start of

research on intelligent sensor systems for interactive behaviors in art scenes, their theory and practice are far from mature and far from practical application needs, especially the high-performance, small volume, and low-cost intelligent sensor systems for mechanical displacement measurement are yet to be further developed.

Based on the research on the structure and performance of traditional inductive displacement sensors and the theory and practice of intelligent sensor systems, this paper designs an intelligent displacement sensor system for live interaction of digital media art, which has the characteristics of small size, low cost, long life, large range, fast response speed, and high intelligence, etc., and obtains satisfactory results in practice. The design of digital art installations requires various technical means such as interactive visual technology, touch technology, physical interactive technology, code programming technology, and virtual/augmented reality technology to realize. In terms of infrastructure, improve the information technology and intelligent facilities and equipment of art and museums, establish the management information system of public cultural space, and improve the information level of infrastructure. In terms of experience, establish fun experience spaces that can interact with audiences, including setting up interactive cultural experience zones for books, paintings, seal engraving, etc., to enhance interactive tours and fun experiences in art and museums. With natural language processing techniques, it is possible to analyze the user's emotions in social media, text input, or voice interactions, and there is already a lot of research available [3]. Multimodal technology is widely applied to the various sensory experiences of art to the users through gestures, touch, gaze, and other forms of multimodal interaction to stimulate various perceptions and responses of art to the users and thus achieve a higher level of experience [4].

II. RELATED WORK

Study [5] says for the user's own experience during multimodal interaction, multiple multimodal modalities with each other for information processing can bring a more efficient and better user experience to the user. Jiang N [6] suggests that the relationship between "immersive scene design" and digital media art is a shadow of one another and that immersive scene design in digital media art is not a new thing that emerged out of nowhere but has undergone a process from quantitative to qualitative change with the progress of technology and exploded in the present. Adams C [7] proposes

the concept of "affective computing", which is a technology that deals with processing, recognizing, and imitating human emotions and feelings. A system with affective computing capabilities can gain insight into the user's emotional state and adjust various factors in human-computer interaction to positively affect the user. Tiwari P [8] explores the progress made in applying affective computing techniques to real-world problems involving human-computer interaction. He points out that effective computing must be developed as a user-centric technology to benefit the global digital economy. In the research [9], the perspectives and applications of emotion in HCI research are discussed in detail, a framework for effective HCI research is proposed, and related methods and techniques are categorized, thus opening up the field of HCI research on human-computer emotional interaction.

In addition, research has given recognition and attention to the application of AI technologies such as decision prediction, image recognition, pattern recognition, and voice interaction in museums, and many studies have explored the means of using AI and other technologies from the perspective of enhancing the overall experience of the audience. The research [10] uses intelligent agent (Agent) systems to assess the level of human-computer interaction through AI to enhance the user experience. By analyzing user interactions in the museum, better interactions and interactive exhibition modules can be designed. On the other hand, Rendell J [11] collects user data based on the viewer's mobile interface, which is used to build a personalized collection recommendation system that brings people and artifacts closer together. Colangelo D [12] points out the correlation between human-computer interaction and the development of artificial intelligence, puts them in a time dimension to sort out their development and evolution, and summarizes the characteristics of interaction design in the era of artificial intelligence. What you think is what you need, what you see is what you see, and what you see is what you get. In addition, under AI, human-computer interaction will break through the boundaries of the graphical user interface and turn its attention to the human emotion and consciousness space for context-aware, consciousness-aware, and emotion-aware computing. The author in [13] proposes three interaction modes: "passive, forced active, and active" between human and intelligent products, and analyzes the differences between them in terms of interaction starting point, initiative, data-driven and AI decision-making to improve the coherence and effectiveness of interaction and help users improve their efficiency and experience. Furthermore, [14] explores the industries and products empowered by machine learning, natural language processing, image processing, robotics, and other technologies in the era of artificial intelligence. Through the case study of smart water purifiers, the innovative, emotional, and information architecture optimization features embodied by the interaction design of artificial intelligence are pointed out. Similarly, [15] points out new ideas that need to be introduced for human-computer interaction-natural

interaction, virtual reality technology for reality-based interaction, augmented reality, context-aware and affective computing, voice interaction, and multimodal interfaces as support. Wohn D Y [16] points out that effective computing is one of the foundations for establishing a harmonious human-computer environment. Since then, most of the research has focused on the field of computer science, trying to improve emotion recognition methods and enhance the accuracy rate by starting from expression, speech and physiological data, and multimodal fusion. Frenneaux R [17] thoroughly analyzed the current emotion models, facial expression interaction, speech signal emotion interaction, physical behavior emotion interaction, physiological signal emotion recognition, emotion extraction from text information and emotion intelligent agents, etc.

III. DESIGN OF INTELLIGENT SENSOR SIGNAL-ASSIST SYSTEM BASED ON INTERACTIVE BEHAVIOR IN THE FIELD

A. Intelligent Sensor System Design Scheme

One of the principles of digital experience design is that it is human-centred. Digital experience design provides a new interaction and expression for the user's experience, further stimulating deeper thinking of the user. Traditional measurement systems cannot store sensor-related information. They cannot be networked, which makes the system not have the function of intelligent identification of sensors and remote monitoring and has failed to meet the interactive behavior of art scene measurement requirements. With the development of modern information technology, intelligent networked measurement systems have come into being. The core technology of an intelligent networked measurement system is intelligent sensor technology, which is also the focus of this paper. Sensor technology, from the birth of the present, has gone through the "traditional sensor", "intelligent sensor", and "networked intelligent sensor" development stages [18]. Networked intelligent sensors of a wide range of different control network standards, access to each other inconvenient, so now the networked intelligent sensors are moving toward standardization.

This paper is designed based on the need for interactive behavior in the digital media art scene, using the intelligent sensor standard of IEEE1451.3. IEEE1451.3 establishes the standard for the interface between analog transmission networks and intelligently networked sensors. IEEE1451.4 establishes the standard for interconnection between intelligently networked sensors in a small space range. IEEE1451.5 establishes that IEEE1451.6 is about CANopen protocol transmitter network interface-related standards. IEEE1451.7 is about RF tag system transmitter communication protocol and transmitter electronic data form format-related standards. Fig. 1 below shows the framework of the standard protocol cluster used in this paper.

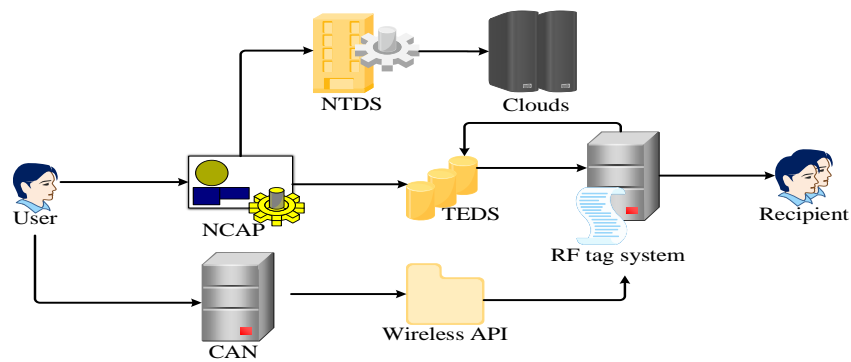


Fig. 1. Standard protocol framework.

The IEEE1451 standard separates the sensor node from the network implementation and is mainly divided into intelligent transmitter modules and network adapter modules. The different standards of IEEE1451.2-7 are used to define different communication protocols between intelligent transmitter modules and network adapter modules. The interface between intelligent transmitter modules and network adapter modules in the IEEE1451.2 standard is called Transmitter Independent Interface TII. The network adapter and intelligent transmitter modules have the same frequency and address arrays stored internally. The network adapter module sends the frequency and address numbers in the array to the intelligent transmitter module in binary form. ACK2 is set high if the element in the frequency and address number arrays is 1 and low if the element is 0. After the Intelligent Transmitter Module is connected to the Transmitter Independent Interface TII, it first communicates with the Network Adapter Module with three handshakes, then receives the binary numbers of frequency and address numbers from the Network Adapter Module, and converts the received binary numbers to decimal to find the corresponding frequency and address for its own IIC configuration [19]. Finally, the network adapter module will perform an IIC communication test with the intelligent transmitter. After the successful test, it will inform the upper computer software that the intelligent transmitter module is connected.

However, with the rapid development of the entire intelligent sensor processing platform and also as a class of typical entry systems, the entire intelligent sensor processing platform is becoming more and more complex, more and more functional requirements of the project, as well as the core microprocessor processing power continues to increase, the intelligent use of interrupts for functional switching of the front and back system defects are increasingly exposed. The first to emerge is a real-time multi-tasking kernel system; such operating systems can provide application programming interface API to the CPU and management of the microprocessor. With the progress of the sensor system intelligence requirements, system applications have become more complex, especially with the rise in communication needs, only the embedded operating system cannot meet. The embedded operating system gradually developed into a complete real-time multi-tasking, multi-functional operating system including file, network, development, and debugging environment, the application and development of IoT-related

technologies, embedded OS, with a complete task (process), scheduling, memory management scheduling and other kernel functions, in addition to this also has the network and other functions of the service capability, reliability, real-time, scalable and reliability, real-time, cut and portability, etc. are also greatly developed [20]. Because of these features, embedded OS has become one of the decisive components of the smart sensor processing platform. freeRTOS is also in the field of smart sensor processing platforms today and occupies a rising rate of emerging RTOS, mainly characterized by small resource consumption, lightweight systems, short development cycles, ease of getting started, completely open-source free, and other features. Networked, especially sensor systems using wireless network technology, also make intelligent sensor processing platforms an important part of IoT devices. In sensor processing platforms that use wireless communication technology, the use of wireless technology enhances the maintainability of the platform. It is even cheaper because it saves costs such as wiring, and the competitiveness rises in the market. The application of wireless communication technology in the field of smart sensors makes up for many shortcomings of smart sensor processing platforms using traditional wired communication methods, breaking through some limitations such as physical environment restrictions on wiring, as well as not being easy to expand, higher costs, and other troubles. The use of wireless network technology in line with the application scenario can expand the smart sensor processing platform in the fields of medical health, environmental monitoring, consumer electronics, and other applications.

IV. APPLICATION OF INTELLIGENT SENSOR SIGNAL-ASSIST SYSTEM IN DIGITAL MEDIA ART INTERACTION

The development of digital experience equipment is a change in the traditional experience mode, it changed the traditional information transfer experience based on audio-visual, and through the use of technology to update the user to receive information and the use of products in the form of the user is no longer satisfied with the experience of a single sense, but more inclined to the "one plus one is greater than two" of the multimodal transport experience. The integration and superposition of multiple senses in the process of digital experience can fully mobilize the user's bodily functions, make the reception of information more interesting, and make the cognition of the experience more real [21]. Typical multimodal digital experience has virtual reality technology, which creates a virtual experience environment to give real cognition to the

human brain. It can directly mobilize the integration of emotional modalities of participants through the basic elements of narrative design law, design, and creation of virtual scenes, real or virtual characters, engaging plot, immersive experience, and real atmosphere in the form of stories so that the experience is repeatedly between abstract and figurative. As a combination of art, technology, and media, digital media art immersive scenography belongs to spatio-temporal art if it is categorized according to the criteria of how the art form exists. The experience is repeatedly spanned between the abstract and the figurative, and feedback is given to the virtual environment through multimodal interaction to enhance the overall sensory experience and make it interesting and intuitive throughout the end of the digital experience.

Digital experience device design integrates a variety of media, including science and technology, experience design, aesthetics, etc. It integrates and restructures new art forms to enrich the expression of art. Through digital and multimedia technologies and forms to enhance the quality of the user's feelings in the experience, the interaction process can enable users to explore their thoughts and behaviors above the basic level of perception to achieve the purpose of integrating emotions and resonance. Digital experience devices can be roughly divided into six technical means, which are interactive visual technology, physical interactive technology, touch interactive technology, virtual augmented reality technology,

mechanical interactive technology, and programming interactive technology [22]. Digital experience device provides new interaction and expression for the user's experience, further stimulates deeper user thinking, strengthens the perception of information, the pursuit of convenience and emotion, and makes the perception effect more towards the real feeling. Digital experience design is different from traditional installation design in that it requires not only the design of physical devices but also software support. The interactive characteristics of digital experience devices are different from traditional art devices and usually require interaction with people. Hardware is the basis for artists and designers to present their ideas, and no matter what kind of artistic expression is used, it should be carried by hardware technology. With the advent of the experience economy and the continuous development of sensory sensing, human body recognition, and tracking technologies, multimodal interaction is gradually used more frequently in creating digital experience design. Traditional physical experience devices have gradually been replaced by digital experience devices, giving users a multimodal interactive experience process consisting of multiple sensory elements that can receive multiple external communication messages and deeper emotional values. The corresponding design strategy is proposed in the design of the museum installation, and the design route of multimodal interaction in digital experience is shown in Fig. 2.

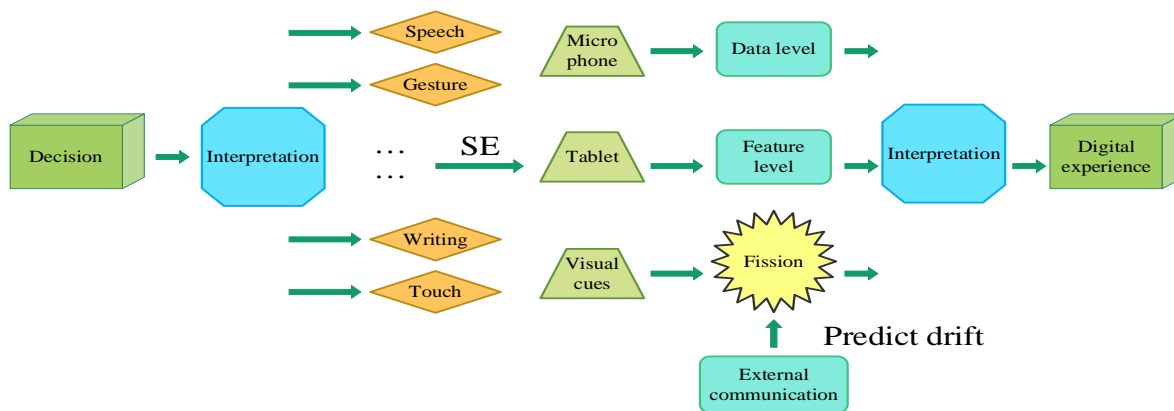


Fig. 2. Design routes for multimodal interaction in digital experiences.

The measurement structure of the capacitive pressure sensor is designed as a differential variable pitch type, consisting of a lower fixed pole plate and an upper half movable pole plate, where the lower pole plate is a silicon substrate covered with a metal oxide film and the upper pole plate is a metal pole plate with a measurement cavity in the middle. The cavity change causes a capacitance change, and the measurement electrode and the driving electrode are led from the two pole plates, respectively. The advantage of using a differential structure is that when the sensor is a non-differential structure, the capacitance value of the capacitor of electrification ($\tau = 1$) is

$$\eta = -\mu \sum_{t=1}^n K_{ij} \Omega_{ij} - K |M_{ij}| \quad (1)$$

When the pole plate is actuated, the plate gap change capacitance is changed $C_y(z)$, i.e.:

$$C_y(z) = C_\mu \cdot \int A \cdot M(x)M(z)dx dz + C_{x0} \quad (2)$$

Then for non-differential structures, the sensitivity is C_y :

$$C_y = C_\mu \cdot \int A \cdot M(x)dx \quad (3)$$

Its nonlinear error b is approximated by:

$$b = (O_k - T_k) \cdot g(k) \quad (4)$$

The comparison shows that using a two-pole differential capacitance sensor is twice as sensitive as a single-pole capacitance sensor and greatly reduces the nonlinear error of the sensor. In addition, the structure also reduces the influence of electrostatic force between the poles and effectively improves the interference caused by environmental factors such as temperature. When the mechanical structure enters the microscopic level, many phenomena that are ignored at the macroscopic level can become factors that cause instability in

the mechanical structure at the microscopic level [23]. For many MEMS devices, the phenomenon of electrostatic absorption is well worth studying. Currently, "workable objects > dynamic visualization > non-workable objects > picture-based static visualization > graphical static visualization > text", from what can be touched to what can be seen in the order of human sensory contact, too abstract, is not popular. Processing involvement and intervention can help users achieve the purpose of human-computer interaction and improve people's experience when they are in contact with digital experience devices. The electrostatic suction phenomenon is the disruption of the dynamic equilibrium of the system when the applied voltage reaches a certain value for the dynamic and fixed poles, and the two poles are sucked together. On the one hand, the existence of the electrostatic suction phenomenon causes instability in measurement. On the other hand, the use of electrostatic force generated by the charge distribution on the pole plates to drive micro devices is also an important research direction.

The control system of the art installation is divided into four modules, including respectively information acquisition module, an information transmission module, a central control module, and a media presentation module, which together build the whole interactive system of the digital art installation and only by combining these four parts can the complete interactive process of the digital art installation be completed. There are many types of information acquisition modules, including simple touch acquisition, sound acquisition, image acquisition, etc., to more complex behavior acquisition, expression acquisition, and other modules. The information transmission module plays a role in the whole digital art installation; when the information acquisition module detects the information, including touch, image, sound, action, etc., it transmits to the central control module through the transmission module, the central control module is also the core part of the whole art installation. It is also the computer's control system, which calculates, processes, and uses the data and programs and then gives them to the media presentation module through the information transmission module. The media presentation module presents the final artistic effect through the media presentation module. The final effect of the media presentation module directly determines the artistic effect of the installation. Using the presentation form of digital media technology can mobilize a variety of senses for the experience, such as using a touch screen to mobilize the sense of touch, allowing the experience to integrate into the art installation or even become part of the installation.

V. A TECHNOLOGICAL APPROACH TO DIGITAL EXPERIENCE IN MULTIMODALITY

Digital experience device design integrates a variety of media, including science and technology, experience design, aesthetics, etc. It integrates and reorganizes new art forms to enrich the expression of art. Through digital and multimedia technologies and forms to enhance the quality of the user's feelings in the experience, the interaction process can make the user explore their thoughts and behaviors above the basic level of perception, to achieve the purpose of integrating emotions and resonance. Digital experience devices can be roughly

divided into six technical means, which are interactive visual technology, physical interactive technology, touch interactive technology, virtual augmented reality technology, mechanical interactive technology, and programming interactive technology [24]. The digital experience device provides new interaction and expression for the user's experience, further stimulates the user to think more deeply, strengthens the human perception of information, the pursuit of convenience and emotionality, and makes the perception effect more towards the real feeling. In emotion computing, the computer does not measure emotions directly. Still, it extracts emotional data by observing external representations of emotions and performs a pattern recognition to infer the current emotional state. Therefore, emotion recognition can be understood as "establishing a mapping relationship between emotional characteristics data and internal emotional states". With the development of augmented reality, mixed reality, and artificial intelligence technology, the future development trend of intelligent sensor-assisted digital media interaction will also be combined and updated with iterations. Emotion modeling is to build a mathematical model to classify and quantify human emotions and then train the model a lot through the emotion database so that the model can recognize various emotions more accurately, so it is the key to emotion recognition.

Interactive cultural creation provides a way of thinking about cultural creation innovation. By integrating cultural and historical knowledge into digital and gamified carriers, users can be inspired by culture through interactive experiences. By injecting the concept of "interactive", we can broaden the form and carrier of cultural creation, change the purpose and thinking of cultural creation development, and upgrade the cultural creation experience, ultimately driving the disruptive innovation of museum cultural creation. The relationship and differences between interactive and traditional cultural creation are shown in Fig. 3

Under the trend of "technology +" and "AI +", the practice of interactive cultural and creative design in museums has taken shape. As a form of cultural creation that focuses on experience, emotional interaction under artificial intelligence is based on the ability to understand and respond to emotional data in interaction, which can precisely give deeper meaning to the cultural creation experience with emotion, thus reflecting the full necessity. Secondly, emotional interaction can gain insight into users' interests, adjust their state, and eliminate frustration; the interactive experience under AI embodies multimodality and initiative and generates natural and real feelings, thus providing more opportunities and possibilities for the interaction design of cultural creation. Through interactive and emotional creative works, the CMA hopes to target the general public with less experience in art appreciation, eliminating their intimidation when facing obscure artworks. An audience survey that examined whether interactive artworks in museums increased engagement showed that 76% of people felt they enhanced their overall museum experience. In comparison, 74% felt it encouraged them to look more closely at the artwork, and those who experienced interactive artworks reported being more willing to delve into cultural history and learn more.

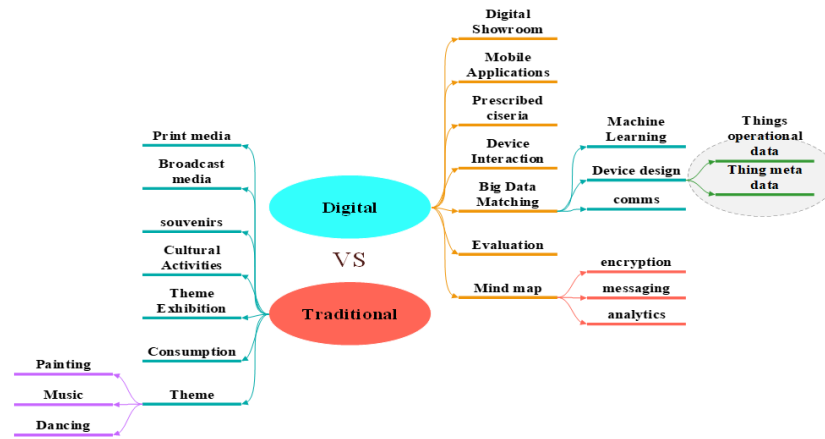


Fig. 3. Relationship and differences between interactive and traditional culture and creativity.

Emotional interaction is the system's mindfulness of human emotions, i.e., the system's mode of understanding human emotions. Emotion computing is the use of artificial intelligence mathematics to calculate human emotions. However, before that, it still requires designers to "specify" emotions in two ways: establishing emotion dimensions and defining key emotions. Establishing an emotion dimension is to describe emotions in a multidimensional and quantitative way [25]. The role of emotion models in emotion recognition is divided into discrete models, dimensional models, and others. The dimensional model can be referred to in emotional interaction design to establish the design dimensions of emotion recognition in performance. Among them, the most recognized and widely used model is the valence-arousal model, based on which the PAD model consisting of pleasure-arousal-dominance is proposed.

Key emotions are scene-related, descriptive words for emotional states, which can be used as typical emotions for emotion recognition and expression of key features. In the emotional interaction design of literature and creativity, first, establish a thesaurus of key emotions; then filter by the degree of emotional granularity (i.e., the degree of detail and complexity of key emotions needed for emotional identification and emotional expression), the greater the granularity, the more specific and clear the description of emotions, and vice versa, the vaguer; it can be described as follows.

$$F(e) = U \omega^e \cdot [K\chi + R\eta]^{-1} \quad (5)$$

Finally, the emotional interaction process is designed based on the key emotions and the stimuli required for emotional expression.

$$F(\gamma) = \sqrt{a^2 + b^2} \cdot c\varpi + \kappa\gamma^d \quad (6)$$

In practical design, key emotions and emotional dimensions can be used in combination. By mapping key emotions to quadrants composed of multiple emotional dimensions, we can ensure that various emotional states are covered and also

facilitate cooperation and communication between designers and developers. Emotional interaction is a new kind of interactive experience, which is very different from the graphical interaction interface represented by smartphones in terms of input and output modalities and elements and the form of emotion recognition and expression. Secondly, the emotional interaction system has artificial intelligence's decision-making and feedback capability. It involves natural interaction modes, such as gestural interaction, somatic interaction, and even brainwave-intentional interaction [26]. Therefore, the factors mentioned above should be reflected in the emotional interaction interface so that emotional acquisition and recognition can be easily perceived and emotional intervention and expression can be easily understood by people.

VI. EXPERIMENTAL VERIFICATION

A. Smart Sensor Calibration Experiment

Using the least-squares method to process the data, the choice of the highest power of the fitted equation also has a significant effect on the accuracy of the pressure sensor. The fitting algorithm was written using Matlab to process the positive and negative travel data of the sensor to generate fitting equations with the highest power of two, three, and four, respectively, and the correlation coefficient of the fitted curve R^2 , which is the degree of similarity between the fitted curve and the actual curve. According to the comparison results, the fitted correlation coefficients are very similar for the highest power of three and four, while the second power fails to meet the requirements. Considering the difficulty of data processing, the highest power of the fitted equation was finally chosen to be three times, and the fitting error was analyzed according to this fitted equation, according to Eq. 7.

$$L = \sum_{(i,j,k) \in D} -\ln \sigma(z_{ij} - z_{ik}) \quad (7)$$

The fitting error of the measured data was calculated using Matlab to obtain the fitting error curve in Fig. 4.

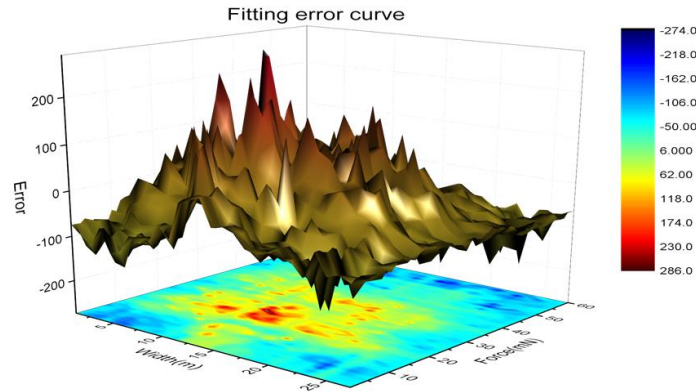


Fig. 4. Fitting error curve.

The resolution is the smallest input signal increment that the sensor is more capable of sensing or monitoring. When the system is running, the sensor to be measured is adjusted to any position after the system is stabilized, and the data from the sensor to be measured is obtained over some time. The installation and measurement process of the smart electrostatic actuated capacitive force sensor is a multi-system fusion installation and measurement process, which theoretically requires a highly stable installation and operating environment. In practice, due to the test conditions, the test bench installation conditions are rudimentary, and the operation process is highly susceptible to vibration, which affects the measurement results and usually takes the form of discarding the data from the previous section. In addition, if there is a deviation in the connection between the sensor to be measured and the displacement element, it will also lead to the elastic deformation of the components, causing abnormal sensor output data. The error is usually reduced by taking the average of multiple data to avoid this situation. Various factors cause noise in the system, and it is also difficult to eliminate, including power quality, component noise, switching noise, system integration of different functional modules cross noise, space magnetic field, capacitance measurement of the noise introduced by the discharge to ground, and the noise in line with the principle of reducing the noise so that it does not affect the measurement results. A filtering circuit is used for each module of the board, the sensor and circuit board are grounded with a metal case, and a twisted shield is used for the

key lines to reduce the noise impact of the above factors. Fig. 5 shows the resolution sampling after correction.

When curve-fitting the sensor output data, the fitted curve does not completely represent the real data, and there is always a certain amount of error present. To reduce the fitting error, the spacing of the collected data is reduced, and the method of fitting higher power terms is improved. In the experiments, fitting polynomials of the highest power three are used to meet the data fitting requirements. The absolute and differential pressure sensors are tested using a portable pressure calibrator. To perform the pressure test, the sensor is connected to the calibrator through a conduit, and the pressure inside it is changed by adjusting the knob of the pressure meter, which in turn leads to changing the pressure value at the detection end of the sensor. Since the interval residual can be regarded as a time-varying detection threshold, the proposed interval residual-based detection algorithm can further address the limitations of the residual evaluation function and a priori threshold in signal processing-based detection methods. Calculating each sensor value can be used as a basis to determine whether the sensor is working properly. Since each sensor value is less than the control limit, the calibration coefficient can be calculated directly using the multiple regression fusion algorithms. Then the data fusion calculation formula is used to calculate Wan. The focus is on the data processing of each sensor. For the sensor, the comparison of the controller display data with the experimentally processed data through simulation resulted in the results shown in Fig. 6.

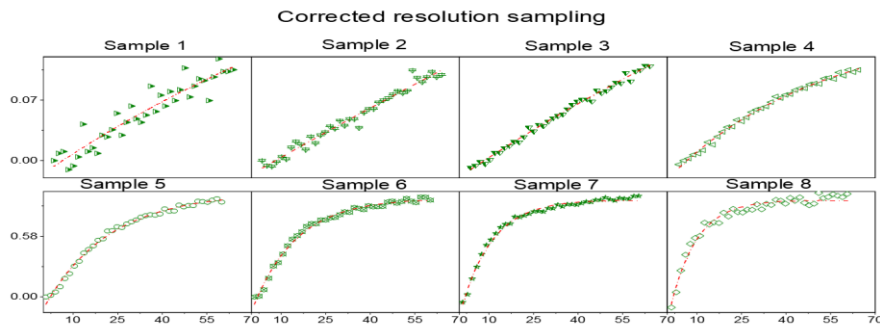


Fig. 5. Corrected resolution sampling.

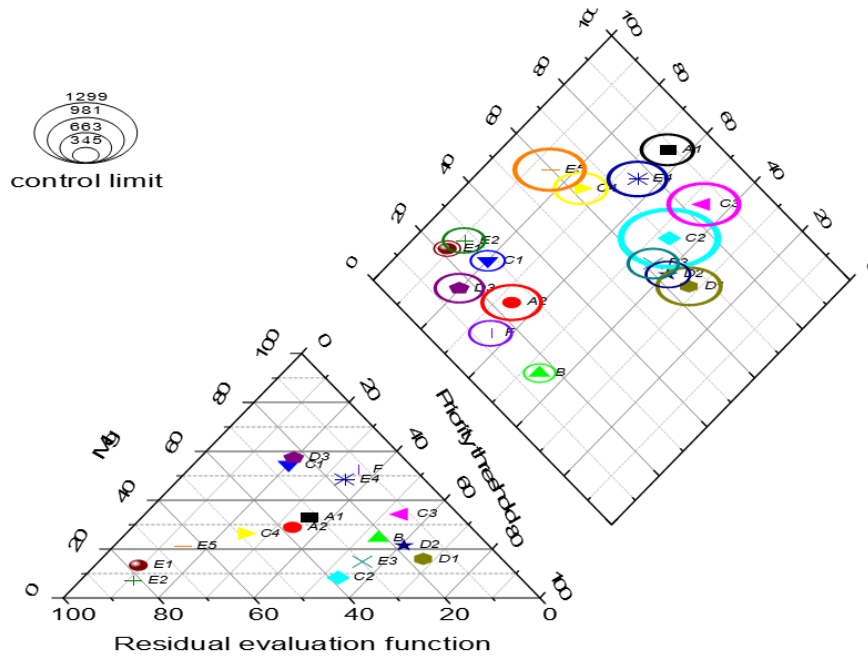


Fig. 6. Sensor residual modulation image.

All components of the common platform should be modularized, including all hardware and software, should use the modular design idea to realize the platform can add or delete components and ensure the least change to other components, mainly to facilitate the secondary developers to be able to customize and tailor the hardware and software according to the actual needs.

B. Smart Sensor-based Digital Media Art Application Field Test

Compared with the traditional design method of creative arts, the intelligent creative arts design implementation process embodies many advantages: First, based on the information collection, opinion analysis, and audience insight capabilities of intelligent sensors, it can more precisely and efficiently locate the emotional interactive creative arts user group, draw a user profile of them, and define the creative arts functions according to the needs and preferences of users. Subsequently, the intelligent design process provides designers with many practical tools. For example, relying on the data brought by museum visitors, AI can help designers filter elements suitable for cultural and creative design from a large number of art resources, which not only makes design elements and design results more in line with users' interests and expectations but also helps expose users to cultural and museum content beyond art exhibits and exhibitions; furthermore, using AI for the content generation to reduce repetitive workload and improve design development efficiency. Based on pattern recognition, AI can evaluate design prototypes and design results, thus becoming a rapid testing tool. In addition, obtaining users' emotional feedback through emotional interactive literature can, in turn, further validate the design and make

improvements in the iteration and serialization of the design. Fig. 7 shows the correlation analysis between emotional performance and interaction behavior.

Fig. 8 shows the construction of the emotional space in which the basic emotions are located. The analysis of the actual data shows that the most dominant affective expressions in affective expressions of the enriched interaction behavior model are positive and agreeable, and the negative and agreeable affective states rarely appear and have no effect on collaboration satisfaction and can be neglected. Therefore, this section focuses on the effect of positive and agreeable effective expressions on collaboration satisfaction in the realistic interaction behavior model. The unstandardized coefficient of the frequency of positive and negative affective performance of the independent variable was 78, significant 18, less than 5, and the constant 86.267, significant 0.000, less than 0.05. The coefficients and constants of the independent variables were significant, using the frequency of positive and negative affective performance as the independent variable for the linear regression of collaboration satisfaction. Emotional interactive creative design by building a library of emotional expression elements so that the system can call all kinds of elements freely according to the results of emotional decision-making. Secondly, by setting the key emotion and emotional dimensions, the intensity and duration of emotional expression of the output elements can be controlled; the scene model is defined to evaluate the effect and impact of the output table elements on the user in that round, and then the emotional interaction system can be adjusted in real-time to ensure that the emotional expression makes the user feel comfortable and natural.

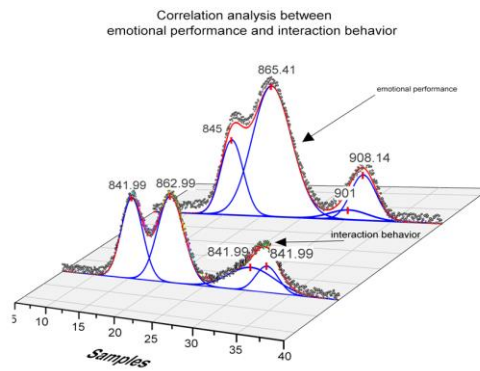


Fig. 7. Correlation analysis between emotional performance and interaction behavior.

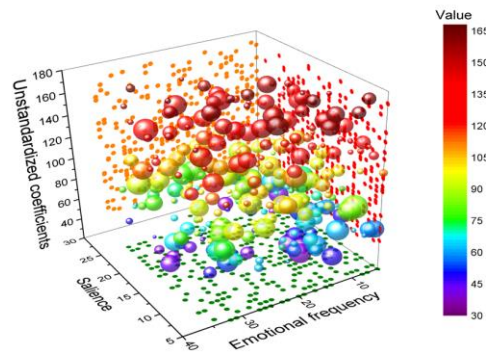


Fig. 8. Construction of emotional space in which basic emotions are located.

VII. CONCLUSION

In terms of theory, the user's feeling consists of multiple sensory elements that can stimulate more profound emotional values during the multimodal interaction experience of receiving multiple externally communicated information. In this paper, according to the multimodal interaction experience process, combined with intelligent sensor signal assistance, the multimodal experience hierarchy in digital experience design is divided into three aspects: basic perception, behavior exploration, and thinking connection. The attack detection and separation method based on state observer is proposed from the cybernetic perspective. The design products are designed from three aspects, including the dynamic physical characteristics of the system, residual evaluation function, a priori threshold design, attack separation under structural vulnerability, etc. The design results include the design of a screen device under basic sensing, the design of a seating device under behavioral interaction, and the design of live interactive behavior of digital media art under thinking connection, the three stages are progressive, further stimulating the user to think more deeply. The three stages are designed to stimulate the deeper thinking of the user and to build emotional resonance so that the user can finally understand the designer's ideas and integrate his understanding of the digital experience. Therefore, the design research on the emotional experience of mobile tour system users also needs to be adjusted according to the current development trend. Relevant design strategies also need to be based on the characteristics of cutting-edge digital media forms, targeted adaptation and output of "technology integration" of user interaction experience digital solutions.

REFERENCES

- [1] K. Gao, H. Wang, J. Nazarko, and G. Chobanov, "Indoor trajectory prediction algorithm based on communication analysis of built-in sensors in mobile terminals," *IEEE Sens J*, vol. 21, no. 22, pp. 25234–25242, 2021.
- [2] G. Shi, X. Shen, F. Xiao and Y. He, "DANTD: A Deep Abnormal Network Traffic Detection Model for Security of Industrial Internet of Things Using High-order Features," *IEEE Internet. Things J*, early access, doi: 10.1109/JIOT.2023.3253777.
- [3] S. Kadry, V. Rajinikanth, N. S. M. Raja, D. Jude Hemanth, N. M. S. Hannon, and A. N. J. Raj, "Evaluation of brain tumor using brain MRI with modified-moth-flame algorithm and Kapur's thresholding: A study," *Evol Intell*, vol. 14, pp. 1053–1063, 2021.
- [4] S. L. Fernandes, U. J. Tanik, V. Rajinikanth, and K. A. Karthik, "A reliable framework for accurate brain image examination and treatment planning based on early diagnosis support for clinicians," *Neural Comput Appl*, vol. 32, pp. 15897–15908, 2020.
- [5] H. Mohammadi Moghadam, A. Mohammadzadeh, R. Hadjiaghaie Vafaie, J. Tavoosi, and M.-H. Khooban, "A type-2 fuzzy control for active/reactive power control and energy storage management," *Transactions of the Institute of Measurement and Control*, vol. 44, no. 5, pp. 1014–1028, 2022.
- [6] N. Jiang and G. Li, "A study of recognising the friction state of revolute pairs based on the motor current signature analysis," *International Journal of Surface Science and Engineering*, vol. 15, no. 2, pp. 87–108, 2021.
- [7] C. Adams, "News on stage: Towards re-configuring journalism through theatre to a public sphere," *Journalism Practice*, vol. 15, no. 8, pp. 1163–1180, 2021.
- [8] P. Tiwari and S. H. Upadhyay, "Advance spectral approach for condition evaluation of rolling element bearings," *ISA Trans*, vol. 103, pp. 366–389, 2020.
- [9] Q. Zhang and K. Negus, "Stages, platforms, streams: The economies and industries of live music after digitalization," *Popular Music and Society*, vol. 44, no. 5, pp. 539–557, 2021.

- [10] M. Pantic, "Gratifications of digital media: what motivates users to consume live blogs," *Media Practice and Education*, vol. 21, no. 2, pp. 148–163, 2020.
- [11] J. Rendell, "Staying in, rocking out: Online live music portal shows during the coronavirus pandemic," *Convergence*, vol. 27, no. 4, pp. 1092–1111, 2021.
- [12] D. Colangelo, "We live here: Media architecture as critical spatial practice," *Space and Culture*, vol. 24, no. 4, pp. 501–516, 2021.
- [13] R. C. King-O'Riain, "'They were having so much fun, so genuinely...': K-pop fan online affect and corroborated authenticity," *New Media Soc*, vol. 23, no. 9, pp. 2820–2838, 2021.
- [14] I. A. Taylor, S. Raine, and C. Hamilton, "COVID-19 and the UK live music industry: A crisis of spatial materiality," *The Journal of Media Art Study and Theory*, vol. 1, no. 2, pp. 219–241, 2020.
- [15] M. D. Thomas, "Digital performances Live-streaming music and the documentation of the creative process," *The future of live music*, pp. 83–96, 2020.
- [16] D. Y. Wohn and G. Freeman, "Audience management practices of live streamers on Twitch," in *ACM International Conference on Interactive Media Experiences*, 2020, pp. 106–116.
- [17] R. Frenneaux and A. Bennett, "A new paradigm of engagement for the socially distanced artist," *Rock Music Studies*, vol. 8, no. 1, pp. 65–75, 2021.
- [18] C.-P. Chen, "Digital gifting in personal brand communities of live-streaming: fostering viewer–streamer–viewer parasocial relationships," *Journal of Marketing Communications*, vol. 27, no. 8, pp. 865–880, 2021.
- [19] T. Portnova, "Art technologization in the context of theatrical science development," *Astra Salvensis-revista de istorie si cultura*, vol. 8, no. Supplement, pp. 701–729, 2020.
- [20] C.-L. Hsu, J. C.-C. Lin, and Y.-F. Miao, "Why are people loyal to live stream channels? The perspectives of uses and gratifications and media richness theories," *Cyberpsychol Behav Soc Netw*, vol. 23, no. 5, pp. 351–356, 2020.
- [21] G. Shi, X. Shen, L. Gu, S. Weng and Y. He, "Multipath Interference Analysis for Low-power RFID-Sensor under Metal Medium Environment," *IEEE Sensors Journal*, 2023.
- [22] D. V. Dunas and S. A. Vartanov, "Emerging digital media culture in Russia: modeling the media consumption of Generation Z," *Journal of Multicultural Discourses*, vol. 15, no. 2, pp. 186–203, 2020.
- [23] C.-I. Park, "A study on the development direction of new media art using virtual reality," *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 21, no. 1, pp. 97–102, 2020.
- [24] Z. T. Chen, "Slice of life in a live and wired masquerade: Playful prosumption as identity work and performance in an identity college Bilibili," *Global Media and China*, vol. 5, no. 3, pp. 319–337, 2020.
- [25] G. He, "Schema interaction visual teaching based on smart classroom environment in art course," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 17, pp. 252–267, 2020.
- [26] E. Timplalexi, "Theatre and Performance Go Massively Online During the COVID-19 Pandemic: Implications and Side Effects," *Homo Virtualis*, vol. 3, no. 2, pp. 43–54, 2020.

Research on Strategic Decision Model of Human Resource Management based on Biological Neural Network

Ke Xu

Cangzhou Normal University, Cangzhou Hebei, 061000, China

Abstract—Human resource management system is an indispensable part of information strategy construction. Based on the theory of biological neural network, this paper constructs the strategic decision model of human resources management, then uses the micro-integration method to predict the demand for human resources, and solves the quantification problem of human resources supply prediction. In the simulation process, the model analyzes the current situation of the personnel management system and the necessity of research and plans and designs a computer-aided personnel management information system based on the Client/Server biological neural network structure. Personnel quality evaluation through the evaluation and analysis of the quality of the evaluated, to provide effective reference information for the enterprise personnel decision and index selection, the enterprise human resources allocation, use, training and development is of great significance. Neural networks rely on the powerful data storage, processing and computing capabilities of computers to help enterprises respond quickly to changes in external market conditions, improve the efficiency of decision-making, and create greater value for enterprises. Through experimental testing, it is found that when the iteration is 5, the network verification results have the best consistency. When the iterations reach 7, the standard of training target error set in this paper is reached. When the samples reached 60, the screening accuracy of the network reached 92.18%; when the samples increased to 80, the screening accuracy was further improved to 92.84%, indicating that the screening accuracy of the network increased with the training samples, which could be used to detect and classify samples quickly, objectively and accurately.

Keywords—Biological neural network; human resources management; strategic decision making; index selection

I. INTRODUCTION

With the gradual improvement of biological neural networks, the use of biological neural networks to establish a comprehensive evaluation system can often achieve unexpected results [1]. Especially for the comprehensive evaluation of those systems with many evaluation objectives and complex relationships between the objectives, the biological neural network model can often achieve better results [2]. There are many mature methods for a systematic, comprehensive evaluation, such as the fuzzy evaluation method, grey system evaluation method, AHP (Analytic Hierarchy Process) and so on [3-5]. Using the analytic hierarchy process to determine the weight can weaken the human factor [6]. Still, AHP requires that the elements in the

hierarchical structure system of indicators are independent of each other [7]. Otherwise, this method cannot be applied. Still, there is often a dependency relationship between these indicators [8].

The key and difficult point of the human resource planning scheme design is the choice of the human resource forecasting method. The development and management of human resources is the essence of labour and personnel management and the core work of enterprise management [9-11]. Human resource planning is the business foundation of human resource development and management. Without reasonable and supporting human resource planning, any plan can only be a piece of paper [12]. Many enterprises lack systematic operation of their human resources, which will inevitably affect the development of human resources and the improvement of labour productivity [13]. Therefore, companies should raise awareness of the importance of human resource planning. By using this system, the personnel department can promote the standardized management of the personnel department and improve the management efficiency and level; it can conveniently and quickly organize and manage the personnel information originally scattered in various departments and provide reliable data for the scientific decision-making of the unit [14].

Based on the biological neural network theory, this paper constructs a strategic decision-making model for human resource management. First, on the basis of analyzing the characteristics of the enterprise and the basic types of human resource strategy, the basic classification of the human resource strategy of the enterprise is determined. Secondly, it extracts four dimensions for the index, uses the method of biological neural network analysis to take indicators of human resources, and then builds the enterprise's human resources strategic decision-making index system. Finally, it takes S company as an example to carry out empirical research. Through comparison, grey forecasting technology and neural network forecasting technology are used to forecast the human resources demand of enterprises. Enterprises can choose one according to their actual situation and then use the micro-integration method to analyze the results and amendments. On this basis, the article applies the current situation verification method and the biological neural network prediction to the enterprise human resources supply forecast, analyzes and demonstrates the applicability. It provides a relatively novel and effective forecasting method for enterprises to carry out human resource planning.

II. RELATED WORK

The focus and difficulty of human resource planning for enterprises is human resource forecasting, and the forecasting method should be selected according to the characteristics of the enterprise [15]. Due to the lack of a relatively mature set of personnel scheduling procedures, it is impossible for enterprises to predict personnel needs reasonably. Therefore, an important part of their human resource planning is establishing a forecast model of enterprise human resource supply and demand [16].

The management system is not combined with the actual needs of human resource management. Still, it is only professional research and development and does not systematically and comprehensively cover human resource management content. Kouhalvandi [17] proposed a dual-objective binary integer programming (BOBIP) model in a fuzzy environment to obtain the best results of person-post matching, calculate the fuzzy utility similarity function to measure the satisfaction of employment outcomes and adopt a mixed integer programming model which realizes two-way matching between enterprises and students. Chen [18] also constructed an evaluation index system with scientific and technological personnel as the research object and proposed a fuzzy comprehensive evaluation model based on ANP (Analytical Hierarchy Process) to realize the matching grade evaluation of scientific and technological talents. Comparatively speaking, this paper summarizes the factors affecting employee dismissal from the three levels of individual, organization and environment, defines the sources of employee dismissal risk, and then combines the characteristics of employees at the beginning of their career with the advantages of sensitivity, relative independence, extensive, measurable, comparable and operable.

In the study of the matching between people and organizations, Falah [19] pointed out that when the ability of the individual is what the organization needs, the matching between the two can be achieved; if the organization can meet all the needs of the individual, it can also promote the matching between the two parties. The situation's impact includes employees' job selection, career planning, etc. Chen [20] believes recruiters are most concerned with selecting candidates with the required knowledge, skills, and attitudes. In contrast, this paper attempts to use the artificial neural network itself has the characteristics of parallel data processing, good fault tolerance, self-adaptation and self-learning, and better nonlinear function, so as to systematically analyze the enterprise personnel quality structure, which is more generic and concise. Based on the general competency model, some scholars combine AHP and fuzzy proximity to calculate the efficiency matrix of each employee and use the assignment model to achieve the best match between all employees and positions in the enterprise [21]. Workflow mainly realizes the transfer of work between the business process participants. Comparatively speaking, this paper selects several early warning indicators of turnover risk to form an early warning index system of turnover risk. Then, on the basis of the status data of early career risk warning indicators, the principal

component analysis method is used to propose the early warning index of turnover risk for employees at the beginning of their career, and the index system is more perfect and streamlined [22-24].

III. CONSTRUCTION OF A STRATEGIC DECISION-MAKING MODEL FOR HUMAN RESOURCE MANAGEMENT BASED ON A BIOLOGICAL NEURAL NETWORK

A. Biological Neural Network Hierarchy

For the enterprise performance evaluation system based on the biological neural network, here is the performance evaluation of the customer representative index of the enterprise customer centre as an example. To evaluate the performance of the customer representative, as long as the performance indicators and relevant data of the customer representative are input into the trained network, the corresponding output, that is, the performance score $s(i, j)$ of the customer representative, can be obtained.

$$\sum p(i, j)/p(i) = 1 \quad (1)$$

$$s(i, j) = \sum a(i)x(i) + w(i)w(j) \quad (2)$$

After the network $w(i)w(j)$ is trained, that is, after the weights of the nervous system are determined and the structure is stable, new data can be processed. The corresponding comprehensive evaluation results can be given. With reference $1-y(t)$ to the set judging standard, the performance $z(x, y)$ of the customer representative can be automatically determined according to the score.

$$z(x, y) \in z(y(t) - x) + z(1 - y(t) + x) \quad (3)$$

$$\sum x(i, m)/x(i) = x(i)/x(m) \quad (4)$$

The neuron transfer function $x(i)$ in the middle layer of the network $x(m)$ adopts the sigmoid tangent function tansig . This is because the output of the function lies in the interval $[0, 1]$, which just meets the requirements of the network output. The training function uses the trainlm function $d(m, mt)$.

$$d(m, mt) = (y(t) - c(t)) * c(mt - 1) - c(1 - m) \quad (5)$$

The network modelling principle $y(t)-c(t)$ is that when the input vector is farther away from the weight vector, the output of the radial base layer is closer to 0, and the output of the linear layer is less affected; and when the difference between the vectors is 0, the output of the radial base layer is 0. For the specific design and implementation of the network, the modelling function in Fig. 1 can be used to output the function.

15 enterprises were randomly selected to form the sample population. The model evaluates the natural distribution of the status of 15 samples under 20 primary selection indicators and, according to the evaluation results, divides them into five levels: very poor, poor, fair, good and very good. The influencing factor of the department is the state of human resources itself, which can be divided into two second-level indicators. The remaining 8 indicators of environmental conditions constitute the primary selection indicator set for enterprise human resources strategic decision-making.

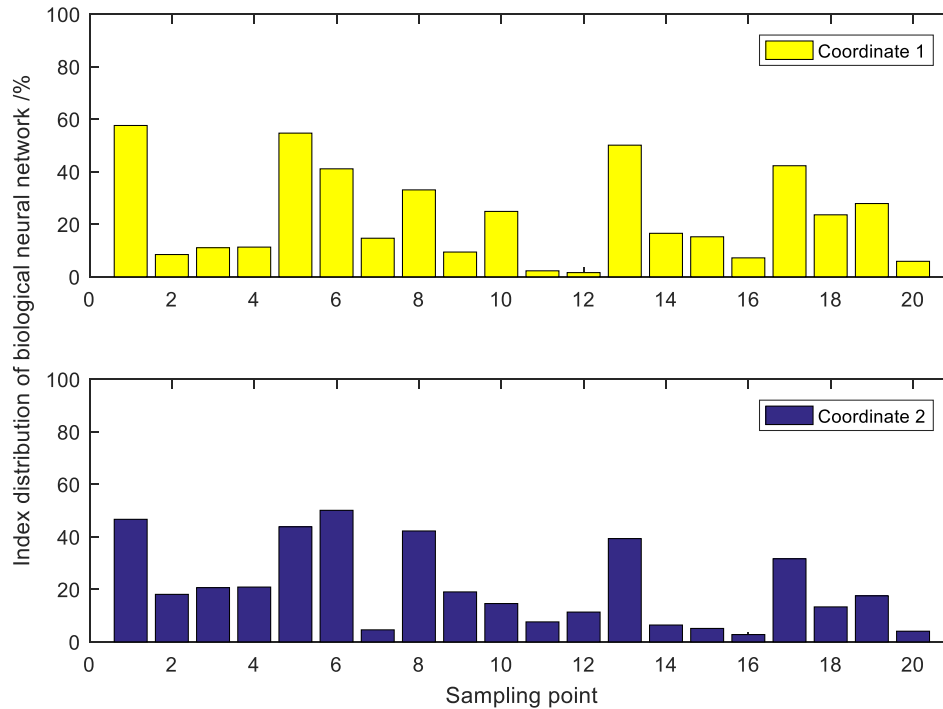


Fig. 1. Distribution of biological neural network indicators.

B. Composition of Human Resource Factors

Through investigation or asking experts to tick the factors of each level of human resource evaluation indicators (important, relatively important, less important, and unimportant), the comprehensive judgment matrix and single-factor evaluation matrix are determined according to the corresponding percentage. Then it can find out their closeness to the fuzzy comprehensive evaluation matrix. According to the principle of choosing the closest, finding the weight vector closest to the comprehensive evaluation matrix is a more realistic weight distribution scheme. When encountering the situation that the solution $b(i)x(i)$ of the fuzzy equation is not unique, there is no better method to select which set of solutions to use as the weight vector $\exp(1-x)$.

$$\sum b(i)x(i) - a(i)x(1 - i) = \exp(1 - x) \quad (6)$$

$$d(x, x') < v(x)t(1 - bx) - v(x')t(1 - ax) \quad (7)$$

In terms of hardware, it means a network computing environment consisting of desktop computers, networks and servers. The meaning of software $d(x, x')$ mainly refers to that a software or application system $t(1-ax)$ is designed as a complex system containing many components, these software components can even be distributed on different machine nodes in the network $\text{rest}(s, t)$, and according to the relative roles of the software components are divided into "Client" and "Server", the client software can request the services of the server software.

$$\text{rest}(s,t) = \{s(i), s(i-1), \dots, s(2), s(1), s/k(s, t^{i-1})\} \quad (8)$$

$$\sum n(i,j)t(i)/t(i-1) = 1 - p(i)/p(j) \quad (9)$$

In the C/S (Client/Server) system, the client $n(i, j)$ always makes a service request to the server first, and the server responds to the client's request $1-p(i)/p(j)$ before sending the service result back to the client. The server never initiates a relationship with the client. The client can also make requests to multiple servers concurrently. In this sense $c(s, t)$, the client is always active, and the server is always passive, so it is asymmetric.

$$c(s, t) \in \left[\frac{s-1}{s}, \frac{i-t+1}{s-1}, i \right] \quad (10)$$

The number of input/output neuron nodes of the network is determined by the external description of the problem. The number of nodes in the input layer corresponds to the number of indicators of the customer representative index of the enterprise customer centre. The text contains the number of customers per customer representative (CSR), the proportion of time that customer representatives work directly for customers, the number of spy orders completed per customer per day, the average time to access customer representatives, and the increase in air talk time per customer. The output result is the evaluation score of the customer representative's performance.

Fig. 2 specially designs the standard unit comparison assignment method to calculate the importance of the alternative indicators. The method is to assign the importance degree of the least important alternative index to 1, use the multiple a of the importance degree obtained by comparing the other indicators with it as the assignment of the importance degree of each index, and the normalized value is used as a method for the alternative index importance shape. It can be

observed that the error meets the requirements. Therefore, the established biological neural network is safe and meets the requirements. This network can be used to infer and predict the total evaluation value of the performance representative according to the existing customer representative performance indicators. In this way, it can be judged whether the

performance of the customer representative has reached the expected target. This method avoids the unreasonable aspects of setting up the weights of various indicators. At the same time, it is scientific and convenient to use and avoids the influence of human factors on the overall performance evaluation.

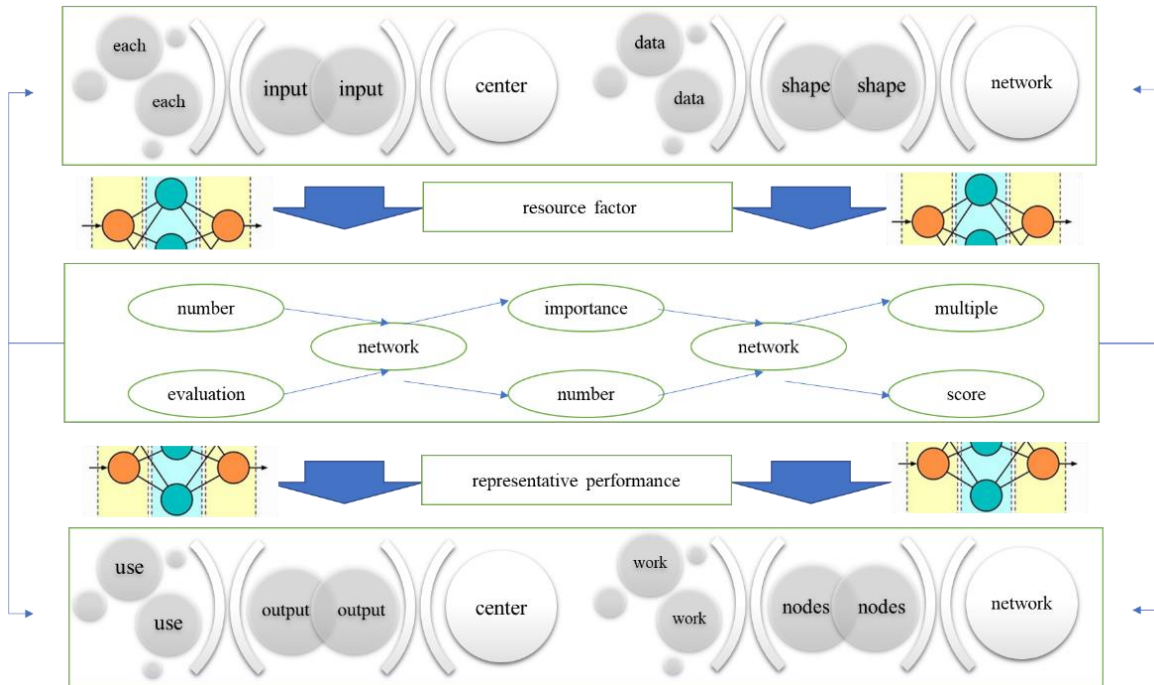


Fig. 2. Human resource factor structure topology.

C. Model Factor Recursion

It can input the model sample matrix data into the MATLAB work window, where the input vector of the training sample is P, the output vector (target value vector) is T, and the input vector of the test sample is PL. After the retraction, a file named P will appear in the workspace window; double-click to open it, and then copy the training sample input value to this file. The MATLAB input matrix takes the employee as the column vector and the factor as the row vector, so in the process of copying, the data s(t, t-1) needs to be transposed before it can be imported into MATLAB so that the matrix e(k)+t of the input vector P is established in MATLAB.

$$s(t, t - 1) = \frac{1}{1 - n} \sum (e(k) + t)^2 + \sum (1 - t)^2 \tag{11}$$

$$\sum_{i=1}^t \left(\frac{Y}{1-r_i} + r - t \right)^2 - \sum \frac{r}{1-r} = 1 \tag{12}$$

The percentage factor 1-r(i) is a method of multiplying the score of each element by the percentage factor that the factor occupies in the total score. Usually, the total score of the four levels of personnel function is expressed by the percentage system, and each element is also evaluated by the percentage system. When measuring, the preliminary score of each factor measured is first multiplied by the percentage coefficient of the factor in the structure. The element score is obtained, and the

scores of each element in the same structure are added to obtain the preliminary score of the structure and then multiplied by the percentage coefficient 1-t of the structure in the total score p(t|1<t) to give the structure score. The accumulation of the four structure scores is the overall score.

$$|(1 - t)p(t|1 < t) + (1 + t)p(t - 1)| < N(t) \tag{13}$$

$$miu(x, y) = miu(x)miu(y) \tag{14}$$

This paper determines the 5-12-1 network structure miu(x, y) through multiple expert experience judgment tests and derivation max(x-t) of empirical formulas, using tansig as the activation function of the network, the number of training samples is 57, and the test data is 16. Then, the training samples are input into the neural network system for training. After 1205 times of learning, a weighted network with a learning accuracy of 0.000001 is obtained. Then the network is called to simulate the test data set, and the neural network simulation results of the test data are obtained. The design work of the system's front end includes interface development and scripting. The development tool adopts powerbuilder6.0. Interface design is the main work of front-end application design, including input and output design. A total of 90 windows and 83 data windows are established in the implementation process.

TABLE I. FACTOR ANALYSIS OF BIOLOGICAL NEURAL NETWORK SAMPLES

Biological neural network codes	Factor analysis text
Public void deletemajorchange();	Each indicator is evaluated
Return salarydao.getbycondition;	The difficulty of obtaining $x - t$
Long majorchangeid;	Independent of each other $ \Delta x - \Delta x t $
Return majorchangedao;	By means of expert scoring
Getbycondition(hql, o);	In the indicator system $rand(m, n)$
Deletemajorchange;	The representativeness $y - t$
Salarydao.deletesalary(salaryid);	The indicators do not intersect
Get.majorchangedao.(majorchangeid);	If the indicators $p(t 1 < t)$
Changegetadd();	The indicator system can only be $N(t)$
Public list { };	Obtaining alternative indicators
This.changegetadd();	At the same level are $k(s, t^{i-1})$
Return majorchangedao;	And simplicity of $1 - p(i)$

Table I divides the difficulty of obtaining alternative indicators into five levels, namely (easy, relatively easy, general, relatively difficult, and difficult), and the corresponding quantitative values are (5%, 25%, 50%, 75%, 95%). The difficulty of obtaining each indicator is evaluated by means of expert scoring. The indicator system's representativeness and simplicity can only be guaranteed if the indicators at the same level are independent of each other and the indicators do not intersect. After the input of the sample, the system learns according to the minimization rule of the mean square error between the expected output and the actual output, and adjusts the weight matrix and threshold vector. When the error is reduced to the required accuracy, the system will stop learning, and the weight matrix and threshold vector will be fixed and become the internal knowledge of the system, which can be called for decision-making or prediction when it is used next time.

D. Optimization of Strategic Decisions

This paper uses neural network modelling to identify strategic decision candidates for specific positions. The sample data adopts the data provided by a talent evaluation research institution. After a four-year investigation, the evaluation and research institution obtained a total of 1,080 sample data, of which 30 were randomly selected for follow-up investigation. The sample data in this paper is based on the actual needs of neural network modelling. After several discussions and consultations with the institution, 25 sample data were selected. One part is used for modelling, and the other part is used for generalization testing. The error ratios of the three test samples are 4.62%, 8.24%, and 6.86%, respectively, and the errors are all below 10%, which does not affect the final strategic type selection.

Since Fig. 3 aims to explore the specific application of neural networks in talent evaluation, a conventional neural network algorithm is used, and the neural network toolbox (Neural Network) can greatly facilitate the network design process, so this paper uses MATLAB neural network toolbox. The degree of fitness for a job can be described by a single output network. Therefore, the number of output layer units is set to 1. The number of input neural units is determined according to the number of influencing factors; 6 units are taken. The algorithm operation platform adopts MATLAB. Among the 25 sets of data, 18 sets are selected as training samples, and 7 sets are used as generalizations.

Similarly, the data matrix of the output vector T can be established: enter T=zeros (5, 20) in the MATLAB command workspace; after the retraction, find the T file in the workspace window and double-click to open it. Unlike the input of P, the T vector matrix must be manually input. Since the MATLAB input matrix uses employees as a column vector, it is necessary to input according to each column, and each employee has a different output value (level). There are different input forms for each level.

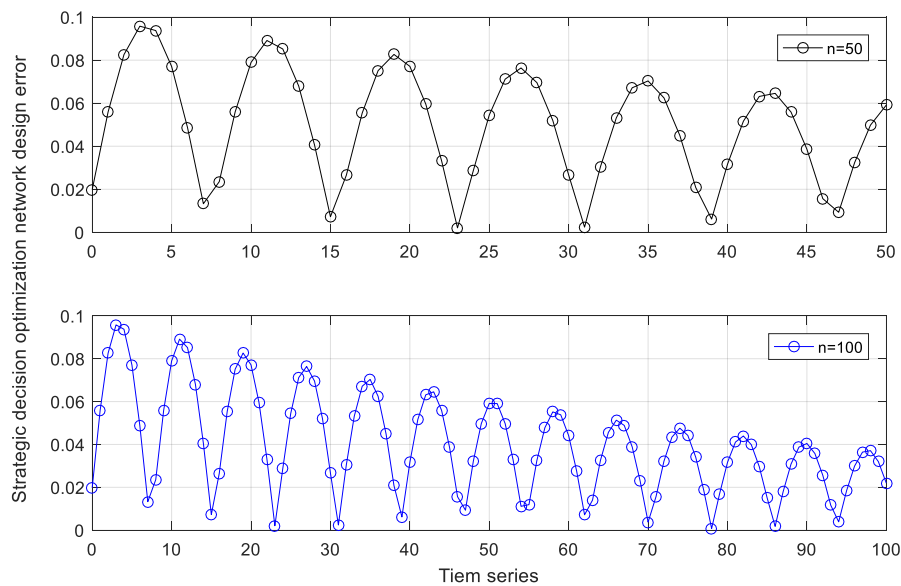


Fig. 3. Strategy decision optimization network design results.

IV. APPLICATION AND ANALYSIS OF HUMAN RESOURCE MANAGEMENT STRATEGIC DECISION MODEL BASED ON BIOLOGICAL NEURAL NETWORK

A. Preprocessing of Biological Neural Network Data

According to the requirements and characteristics of the biological neural network, after the test, feedback adjustments were made according to the test results, and necessary modifications and additions were made to the indicator system. Various quality content is tested separately to obtain closer to the real evaluation results. 120 samples collected in this paper were distributed to A1, simulating the situation in which enterprises were subjected to many CVs (curriculum vitae). After all, samples are marked and numbered in Fig. 4, the top 20 (screening rate $P=20\%$) best talent resumes are screened out using the traditional resume screening method with reference to the weights of each indicator determined by the business owner in the step.

The goal of the early-career employee dismissal risk early warning system is to realize the evaluation of employee dismissal risk status on the basis of the collection of employee dismissal risk information, output the dismissal risk early warning signal, clarify the source of risk, propose risk treatment countermeasures, and achieve the elimination of risk status or take countermeasures to reduce the loss caused by employee dismissal. At the same time, it can be clearly seen that although samples 1 and 2 are in the same matching level, sample 2 has a higher degree of membership than sample 1 for the evaluation level of person-post matching, so it can be judged that sample 2 has a higher degree of membership. The matching degree is better than the No. 1 sample. According to the evaluation results of the matching degree of people and positions of the samples, it can be known that samples No. 1-2 show a good degree of matching. In contrast, the matching degree of sample No. 2 and the position of the project director is not good. The human resources management department can consider implementing it when necessary. The input process

should be as simple and clear as possible to reduce the occurrence of errors, and the input system should have fault tolerance and data verification functions. For example, in the input interface, the basic information of employees is concentrated in one interface, which can not only avoid repeated input of information, reduce the input work scene, but also maintain the consistency of data; use the selection method to standardize the content to make it simple and convenient to fill in, use the Tab control to achieve file cabinet entry, the Tab control can organize a large amount of information or controls in a small space. Using the Tab control and DataWindow filtering function, a large amount of information is concentrated in one window, which is convenient for user management and more efficient than the general method.

From the experimental results in Fig. 5, it can see that the screening method based on a biological neural network is superior to the traditional screening method from the time dimension. After completing all the above steps, MATLAB is used to build a biological neural network, randomly select 2/3 (80) samples from the qualified samples and unqualified samples for training, and the remaining 1/3 (40) samples are used for training. Screening 24 resumes from 120 resumes, the traditional screening method takes nearly 50 minutes. Under the same experimental conditions, the resume screening method using the biological neural network only takes 1 to 2 minutes. In the actual medium and large enterprises or talent market, talents are often screened from thousands or even more resumes. Therefore, with the increase in the number of resumes, the method in this paper can save double the time cost. To sum up, when predicting, it is necessary to determine a smooth factor of an appropriate size so that the dependent variables of all training samples can be fully learned and the distance between different training sample points and prediction samples is considered. The smaller the distance between the samples, the larger the corresponding weight of the dependent variable, which in turn helps the network to make better generalization predictions for new samples.

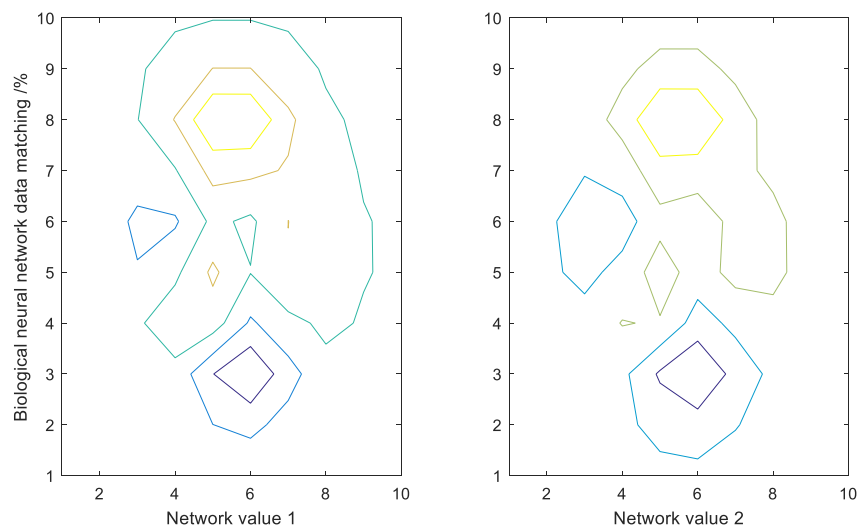


Fig. 4. Biological neural network data matching.

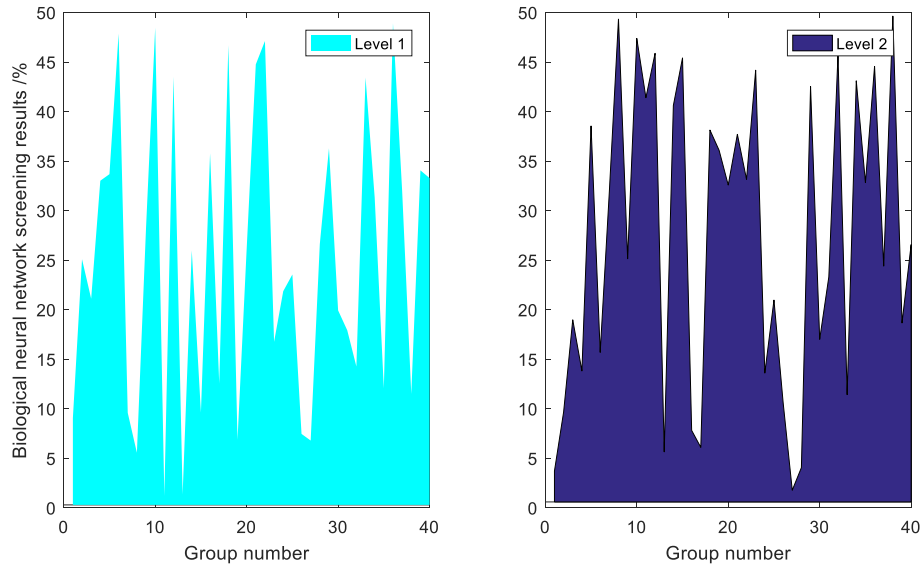


Fig. 5. Screening results of biological neural network.

B. Human Resource Preparation

There are many factors that affect the value of human resources establishment, and which indicators are used to evaluate human resources have a great impact on the accuracy of the results. In this example, seven-factor indicators are used to evaluate the value of human resources, some of which also include secondary indicators, and the seven categories of indicators are selected by the method of clustering, which is scientific to a certain extent. But there are also some defects, such as salary income and other factors that affect the value of human resources are still not included. Therefore, the establishment of a complete set of indicators and factors that can fully reflect the value of human resources is crucial for scientifically and rationally evaluating the value of human resources. By designing and distributing questionnaires (resumes), this paper finally obtained a total of 123 sample data. After preliminary sorting and screening of the data, 3 samples that did not meet the requirements were eliminated.

All data are placed in the whole network. Fig. 6 performs linear regression on the actual output of the network and the corresponding expected output. It is found that the overall output value tracks the expected value better, and the corresponding $R=0.97822$, which is very close to 1. Spring realizes the specific business logic processing work and the scheduling and distribution of the processing results. The presentation layer is responsible for the encapsulation and transmission of the request object and the display of the feedback results, and the specific interaction with the database is processed through the persistence layer objects. Although the templates in Spring implement the encapsulation of some database operations, the specific processing is still implemented by Hibernate. After Hibernate processes the data, the results are handed over to Spring Business processing for processing. The interaction between the framework and the database is done through Hibernate, which is the persistence layer. For the operation request from the presentation layer, first, submit the request to the business logic layer where Spring is located for processing.

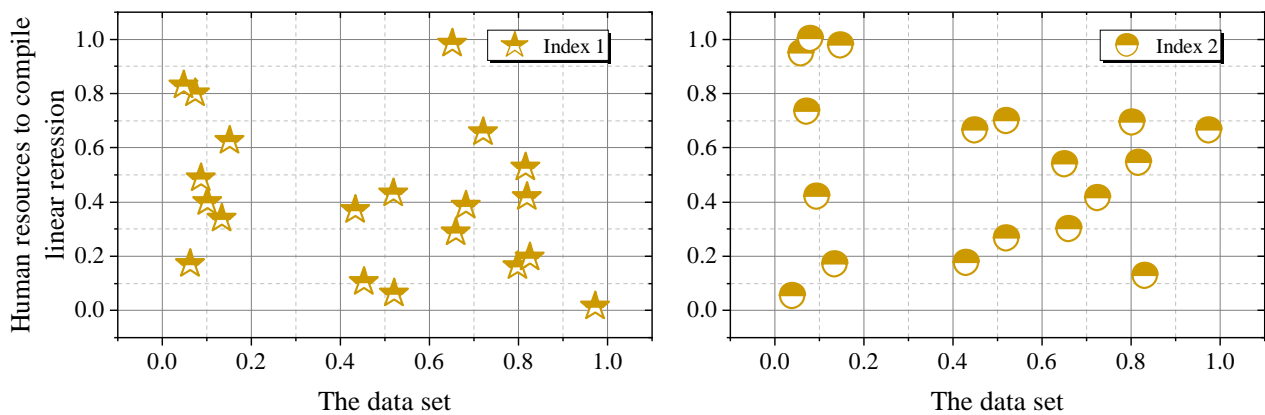


Fig. 6. Linear regression of human resource staffing.

C. Strategic Decision Simulation

This strategic decision evaluation project has adopted various means such as knowledge examination, psychological quality test, interview, simulated performance technology and other means to conduct a comprehensive quality inspection of the assesees. After selecting suitable personnel for the test, feedback adjustment was made according to the test results, and necessary modifications and supplements were made to the index system. Fig. 7 is to organize the data of this evaluation according to the requirements of the aforementioned neural network model and divides it into two parts, the neural network quality evaluation model learning sample set and test sample set, and conduct empirical analysis. It is hoped that the network model can pass the learning to absorb the judgment experience of experts.

The size of the network is the most critical. Usually, the number of samples is at least more than the number of network connection weights and generally requires more than 5 to 10 times. Especially for a three-layer biological neural network, the number of samples must be greater. Otherwise, the network must have redundant nodes, and the systematic error is independent of the characteristics of the training samples and tends to be 0; the network has no generalization ability at all. The value of the smoothing factor is very important and has a great impact on the approximation accuracy and prediction effect of the generalized regression network. If x is very large, y is close to the mean of all training sample dependent variables; conversely, if x is close to 0, y is close to the value of all training sample dependent variables. Therefore, when the sample to be predicted happens to be within the learning range of the training sample, the calculated predicted value will be very close to the expected output. Still, once a new predicted sample is not included in the learning range of the training sample. Then the prediction effect will be significantly reduced, and the network's generalization ability will be significantly reduced.

D. Example Application and Analysis

This paper uses the biological neural network optimization algorithm, the hyperbolic tangent Sigmoid function \tanh is used as the transfer function between the input layer and the hidden layer, the purelin linear function is used as the function between the hidden layer and the output layer, and the trainlm

function is used as the training function, set the error to 0.0001. After three iterations, the network error reaches an acceptable range, and the established network tends to be stable. Finally, the evaluation indicators of the group personnel quality evaluation project were generally determined as major items, and each item was divided into various quality contents for testing respectively, in order to obtain the evaluation results closest to the truth. In the Network/Data Manage window, click it to train the network, select the input vector p from the "Inputs" drop-down list, select the target vector from the "Targets" drop-down list, and train in the training parameter settings in Fig. 8. The number of steps "epochs" is 50, the training goal "goal" is 0.01, and the rest are default items.

The integration between the WebWork presentation layer and the control layer Spring is mainly carried out through two steps: the first step is to initialize Spring when WebWork is used; the second step is to configure Spring while configuring WebWork. All business processing or action processing is unified by Spring. The core of the Spring framework is dependency injection. The implementation of dependency injection in this system includes aspects: injecting data source management and transaction management and injecting interface implementation classes. Through the injection of various interfaces, the unified database operation management of all action processing by transaction management can be realized. The implementation of specific dependency injection is achieved by configuring actions in xml. After obtaining the sample data, the "factor scoring method" is used to evaluate and score each sample. The evaluation results have only two cases, namely 1 or 0. 1 means that the sample is qualified, it means that it has entered the interview process through the audition stage, and 0 means that the sample is eliminated. Compared with the reference [25-27], the artificial neural network used in this paper has a better nonlinear function. This quality evaluation project adopts various means such as knowledge examination, psychological quality test, interview and simulation performance technology to conduct a comprehensive quality investigation of the interviewees. In the process of assessment, the data are sorted according to the requirements of the aforementioned neural network model and divided into two parts as the learning sample set and the test sample set of the neural network quality assessment model for empirical analysis.

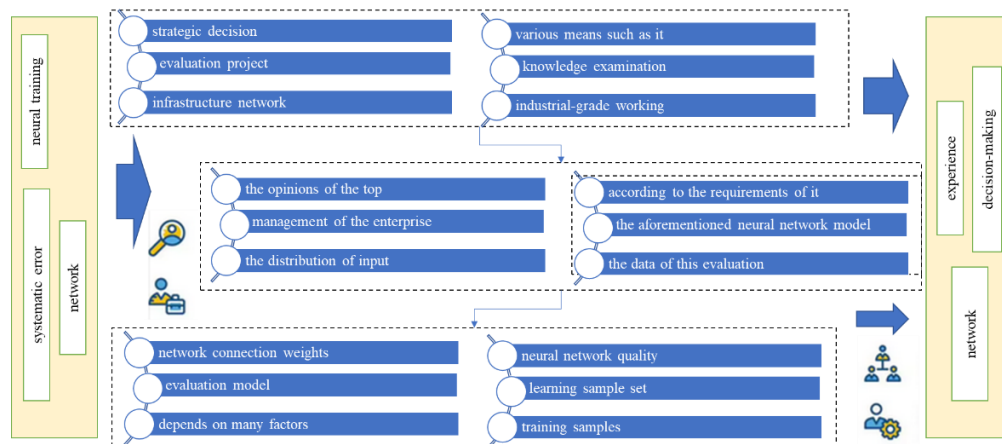


Fig. 7. Strategic decision-making neural network training.

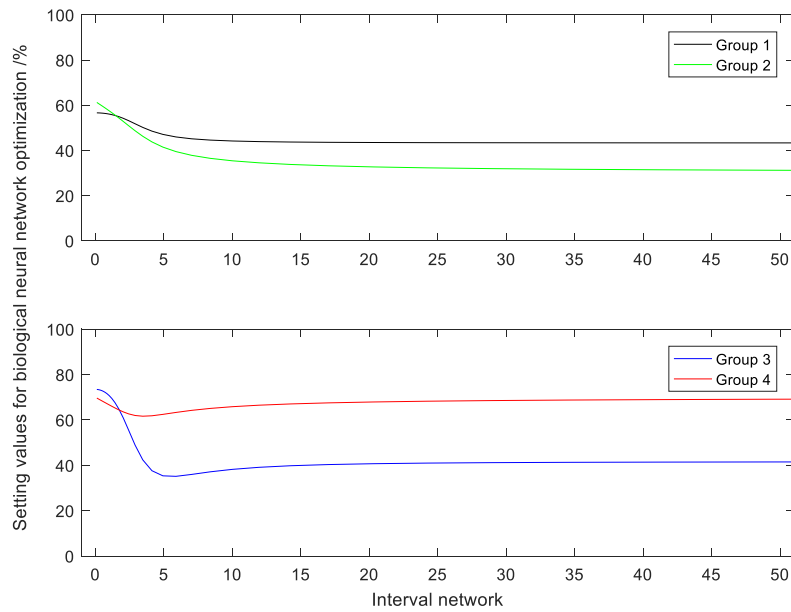


Fig. 8. Biological neural network optimization settings.

The errors of the network simulation results in Fig. 9 are controlled below 5%, and a good evaluation effect has been achieved. The sample evaluation is carried out by the method of "expert evaluation". The model finds 2 entrepreneurs in the field, comprehensively considers the pros and cons of each sample according to the weight distribution, and evaluates all samples as "qualified" or "eliminated". In order to further simulate the real scene of enterprise recruitment, it assumes that the screening rate is $P=0.2$ (Enterprises can set the parameters of P by themselves). Companies need to select the best top 20% of resumes from the 120-point resumes. In the end, the top 24 best samples and 96 less excellent samples were selected. After testing, it is found that the trained biological neural network can detect and classify samples quickly, objectively and accurately. This shows that the neural network method can fully absorb the judgment experience of experts and make more accurate judgments on the test data, which

confirms the availability and accuracy of the biological neural network method.

E. Discussions

According to the analysis of neural network, it can be clearly seen that in the application of BP neural network model in the company, although the output value of BP network is relatively low, on the whole, it fluctuates slightly between the satisfactory value, and the stability is relatively high [28-30]. This paper believes that the decision requirements are met. From the perspective of the prediction effect of the test set, the training effect and accuracy of the improved generalized regression neural network proposed in this paper are higher than that of the generalized regression neural network model, which proves the feasibility of the neural network model proposed in this paper for man-post matching evaluation.

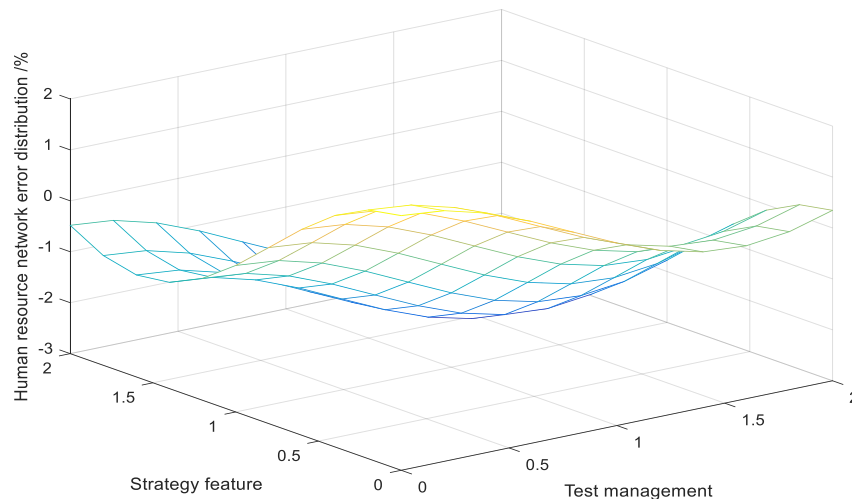


Fig. 9. Error distribution of human resource management strategy network.

V. CONCLUSION

In this paper, a strategic decision-making model of human resource management based on a biological neural network is constructed. On the basis of planning the organizational structure of the enterprise, an enterprise human resource planning scheme is designed. First of all, the model starts from the function of the personnel management system, studies various talent evaluation methods, and tests the learning ability, adaptive ability and function approximation ability of biological neural networks. Secondly, a competency-based enterprise project director personnel-post matching evaluation index system was established, including 5 first-level indicators and 8 second-level indicators of knowledge skills, goal planning, plan promotion, social roles, attitudes and values. In the process of resource management, the process is not standardized, and the degree of information sharing is low. Through the analysis, design, development and implementation of the needs of the target unit, the human resource management system suitable for the company is completed. Finally, an example uses the generalized regression neural network model to conduct a comprehensive evaluation and empirical research on the matching of enterprise project directors and positions. It uses the MATLAB toolkit to complete the specific application of biological neural network modelling in talent evaluation and achieves good results. The training (learning) of biological neural network depends on the original data, which means that the accuracy of neural network simulation is also limited by the accuracy of the original training data, resulting in a significant decrease in the training efficiency of neural network and the accuracy of simulation results. To solve this problem, the method of anonymous participation is adopted, the cost-benefit principle is considered, and some weights are assigned respectively, so as to obtain the final weighted average score of each employee. To a certain extent, the scheme avoids the influence of internal personnel's personal bias on the accuracy of the original training data, and is more conducive to the training of neural networks.

REFERENCES

- [1] Y. Qamar, R. K. Agrawal, T. A. Samad, and C. J. C. Jabbour, "When technology meets people: the interplay of artificial intelligence and human resource management," *Journal of Enterprise Information Management*, vol. 34, no. 5, pp. 1339–1370, 2021.
- [2] A. Mozo, B. Ordozgoiti, and S. Gomez-Canaval, "Forecasting short-term data center network traffic load with convolutional neural networks," *PLoS One*, vol. 13, no. 2, p. e0191939, 2018.
- [3] X. Chen et al., "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Trans Wirel Commun*, vol. 19, no. 4, pp. 2268–2281, 2020.
- [4] P. Tambe, P. Cappelli, and V. Yakubovich, "Artificial intelligence in human resources management: Challenges and a path forward," *Calif Manage Rev*, vol. 61, no. 4, pp. 15–42, 2019.
- [5] Q. Jia, Y. Guo, R. Li, Y. Li, and Y. Chen, "A conceptual artificial intelligence application framework in human resource management," 2018.
- [6] M. H. Saputra and H. S. Lee, "Prediction of land use and land cover changes for north sumatra, indonesia, using an artificial-neural-network-based cellular automaton," *Sustainability*, vol. 11, no. 11, p. 3024, 2019.
- [7] D. Zeng, L. Gu, S. Pan, J. Cai, and S. Guo, "Resource management at the network edge: A deep reinforcement learning approach," *IEEE Netw*, vol. 33, no. 3, pp. 26–33, 2019.
- [8] Y. Hu, Q. Zhang, Y. Zhang, and H. Yan, "A deep convolution neural network method for land cover mapping: A case study of Qinhuangdao, China," *Remote Sens (Basel)*, vol. 10, no. 12, p. 2053, 2018.
- [9] Z. M. Yaseen, M. Fu, C. Wang, W. H. M. W. Mohtar, R. C. Deo, and A. El-Shafie, "Application of the hybrid artificial neural network coupled with rolling mechanism and grey model algorithms for streamflow forecasting over multiple time horizons," *Water Resources Management*, vol. 32, pp. 1883–1899, 2018.
- [10] F. Hussain, S. A. Hassan, R. Hussain, and E. Hossain, "Machine learning for resource management in cellular and IoT networks: Potentials, current solutions, and open challenges," *IEEE communications surveys & tutorials*, vol. 22, no. 2, pp. 1251–1275, 2020.
- [11] S. G. Meshram, M. A. Ghorbani, S. Shamshirband, V. Karimi, and C. Meshram, "River flow prediction using hybrid PSO-GSA algorithm based on feed-forward neural network," *Soft comput*, vol. 23, pp. 10429–10438, 2019.
- [12] S. Yang, D. Yang, J. Chen, and B. Zhao, "Real-time reservoir operation using recurrent neural networks and inflow forecast from a distributed hydrological model," *J Hydrol (Amst)*, vol. 579, p. 124229, 2019.
- [13] B. Hmoud and V. Laszlo, "Will artificial intelligence take over human resources recruitment and selection," *Network Intelligence Studies*, vol. 7, no. 13, pp. 21–30, 2019.
- [14] Y. Chen et al., "Evaluation efficiency of hybrid deep learning algorithms with neural network decision tree and boosting methods for predicting groundwater potential," *Geocarto Int*, vol. 37, no. 19, pp. 5564–5584, 2022.
- [15] Y. Hua, R. Li, Z. Zhao, X. Chen, and H. Zhang, "GAN-powered deep distributional reinforcement learning for resource management in network slicing," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 2, pp. 334–349, 2019.
- [16] C. Qin, H. Zhu and T. Xu, "Enhancing person-job fit for talent recruitment: An ability-aware neural network approach," the 41st international ACM SIGIR conference on research & development in information retrieval, pp. 25–34, 2018.
- [17] L. Kouhalvandi, I. Shayea, S. Ozoguz, and H. Mohamad, "Overview of evolutionary algorithms and neural networks for modern mobile communication," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 9, p. e4579, 2022.
- [18] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019.
- [19] F. Falah, O. Rahmati, M. Rostami and E. Ahmadisharaf, "Artificial neural networks for flood susceptibility mapping in data-scarce urban areas," *Spatial modeling in GIS and R for Earth and Environmental Sciences*, pp. 323–336, 2019.
- [20] D. Xuan, D. Zhu, and W. Xu, "The teaching pattern of law majors using artificial intelligence and deep neural network under educational psychology," *Front Psychol*, vol. 12, pp. 711520, 2021.
- [21] G. Shi, X. Shen and Y. He, "Passive Wireless Detection for Ammonia Based on 2.4 GHz Square Carbon Nanotube-loaded Chipless RFID-inspired Tag," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 9510812, 2023.
- [22] H. Elrehail, I. Harazneh, M. Abuhjeeleh, A. Alzghoul, S. Alnajdawi, and H. M. H. Ibrahim, "Employee satisfaction, human resource management practices and competitive advantage: The case of Northern Cyprus," *European Journal of Management and Business Economics*, vol. 29, no. 2, pp. 125–149, 2019.
- [23] G. Shi, X. Shen and L. Gu, "Multipath Interference Analysis for Low-power RFID-Sensor under metal medium environment," *IEEE Sensors Journal*, 2023.
- [24] F. Cherif, "The role of human resource management practices and employee job satisfaction in predicting organizational commitment in Saudi Arabian banking sector," *International Journal of Sociology and Social Policy*, vol. 40, no. 7/8, pp. 529–541, 2020.
- [25] B. Koziel, A. Ackie A B, and E. Khatib R, "SIKE'd up: Fast hardware architectures for supersingular isogeny key encapsulation," *IEEE*

- Transactions on Circuits and Systems I: Regular Papers, vol. 7, no. 12, pp. 4842-4854, 2020.
- [26] M. Anastasova, R. Azarderakhsh, and M. Kermani, "Fast strategies for the implementation of SIKE round 3 on ARM Cortex-M4," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 68, no. 10, pp. 4129-4141, 2021.
- [27] M. Kermani, and R. Azarderakhsh, "Reliable architecture-oblivious error detection schemes for secure cryptographic GCM structures," IEEE Transactions on Reliability, vol. 68, no. 4, pp. 1347-1355, 2018.
- [28] M. Mozaffari-Kermani, R. Azarderakhsh R, and A. Aghaie, "Reliable and error detection architectures of Pomaranch for false-alarm-sensitive cryptographic applications," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 23, no. 12, pp. 2804-2812, 2015.
- [29] S. Pan, R. Azarderakhsh and M. Kermani, "Low-Latency Digit-Serial Systolic Double Basis Multiplier over $\text{GF}(2^m)$ Using Subquadratic Toeplitz Matrix-Vector Product Approach," IEEE Transactions on Computers, vol. 63, no. 5, pp. 1169-1181, 2012.
- [30] G. Shi, X. Shen and F. Xiao, "DANTD: A deep abnormal network traffic detection model for security of industrial internet of things using high-order features," IEEE Internet of Things Journal, 2023.

Multimodal Deep Learning Approach for Real-Time Sentiment Analysis in Video Streaming

Tejashwini S. G¹, Aradhana D²

Research Scholar, VTU Belagavi, Karnataka, India¹

Department of Computer Science and Engineering, Ballari Institute of Technology and Management,
Ballari, Karnataka, India^{1,2}

Abstract—Recognizing emotions from visual data, like images and videos, presents a daunting challenge due to the intricacy of visual information and the subjective nature of human emotions. Over the years, deep learning has showcased remarkable success in diverse computer vision tasks, including sentiment classification. This paper introduces a novel multi-view deep learning framework for emotion recognition from visual data. Leveraging Convolutional Neural Networks (CNNs) this framework extracts features from visual data to enhance sentiment classification accuracy. Additionally, we enhance the deep learning model through cutting-edge techniques like transfer learning to bolster its generalization capabilities. Furthermore, we develop an efficient deep learning classification algorithm, effectively categorizing visual sentiments based on the extracted features. To assess its performance, we compare our proposed model with state-of-the-art machine learning methods in terms of classification accuracy, training time, and processing speed. The experimental results unequivocally demonstrate the superiority of our framework, showcasing higher classification accuracy, faster training times, and improved processing speed compared to existing methods. This multi-view deep learning approach marks a significant stride in emotion recognition from visual data and holds the potential for various real-world applications, such as social media sentiment analysis and automated video content analysis.

Keywords—Deep learning; emotion recognition; feature extraction; machine learning; sentiment analysis; visual data

I. INTRODUCTION

Emotion recognition from visual data sets, encompassing images and videos, has emerged as a complex and captivating challenge that has garnered increasing attention from computer vision and machine learning. The accurate classification of emotions based on visual cues holds the potential for a multitude of practical applications in the real world, such as social media sentiment analysis, targeted advertising, and automated video content analysis [1]. Deep learning techniques, particularly Convolutional Neural Networks (CNNs) and Re-current Neural Networks (RNNs) have showcased remarkable prowess in various computer vision tasks, including sentiment classification [2]. These advancements in deep learning have opened new avenues for tackling the intricate task of emotion recognition from visual data, fueling optimism for its transformative impact across diverse industries and domains.

The extraction of features from visual data relies on deep learning architectures, particularly Convolutional Neural

Networks (CNNs), which scan images or videos to identify patterns as shown in Fig. 1. These CNNs consist of multiple layers, each responsible for extracting distinct features from the input. Basic features like edges or lines are captured in the initial layers, while higher layers discern more intricate and abstract features associated with diverse objects or emotions [3]. Subsequently, these extracted features undergo classification through a trained model designed to recognize patterns and make predictions. The feature extraction process is iterative and automatable, empowering the CNN to adapt and learn from novel visual data, resulting in improved accuracy and efficiency for visual tasks, including sentiment classification [4]. While current deep learning methods have shown promise in emotion recognition from visual data, they have limitations. These limitations encompass a range of factors that collectively impact these models' overall performance and usability. One significant drawback is their limited generalization ability across diverse datasets and real-world scenarios. Emotions can be expressed in various ways across different cultures, contexts, and individuals, making it challenging for deep learning models to capture and interpret these nuances consistently and effectively. The quest for higher classification accuracy remains ongoing. While deep learning models have demonstrated substantial progress in recognizing basic emotions like happiness and sadness, they often struggle with more complex emotional states that involve subtle variations in facial expressions, body language, and contextual cues. This deficiency in accurately deciphering nuanced emotions impacts the overall reliability of these models, particularly in applications where precise emotional understanding is paramount.

Many deep learning architectures demand extensive computational resources and time for training, which can be impractical for Real-Time or resource-constrained applications. Additionally, the need for vast amounts of annotated data for training can become a bottleneck, as obtaining accurately labeled emotional datasets on a large scale is a resource-intensive and time-consuming endeavor. As such, there is a persistent demand for developing more efficient and effective deep learning techniques tailored explicitly for emotion recognition from visual data. Addressing these limitations requires innovative approaches that focus on enhancing generalization capabilities, refining accuracy across diverse emotional spectra, and streamlining training processes. By harnessing the potential of deep learning while mitigating these constraints, researchers and practitioners can usher in a new era of emotionally intelligent

technologies that better understand and respond to human emotions across various applications.

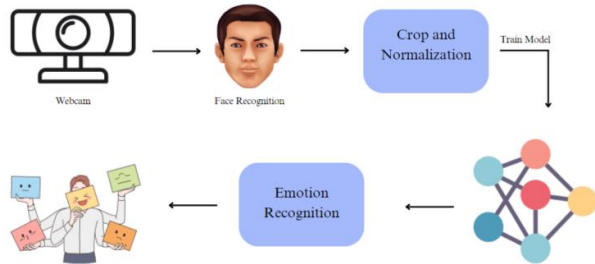


Fig. 1. Block diagram representation of sentimental analysis employing deep learning approach.

The proposed framework seeks to address the limitations encountered in current deep learning methods and enhance sentiment classification accuracy. Our approach involves the integration of CNN-based deep learning architecture to extract features from visual data sets, harnessing their unique strengths to complement each other. We enhance the deep learning model using advanced techniques, including transfer learning, to enhance its ability to generalize effectively. By combining these strategies, our framework aims to significantly improve sentiment classification accuracy, paving the way for more reliable and efficient sentiment analysis from visual data. In addition to the architectural enhancements proposed for improving sentiment classification accuracy, another crucial aspect that our framework addresses is the issue of data diversity and bias. Emotion recognition models heavily rely on the availability of diverse and representative datasets to ensure robust performance across various demographic groups and cultural contexts. However, many existing datasets used for emotion recognition may exhibit biases in terms of under representation or misrepresentation of certain emotions or demographic groups. To mitigate this challenge, our framework emphasizes the importance of curating and maintaining well-balanced datasets encompassing a wide spectrum of emotions and demographic characteristics. By training the CNN-based model on such comprehensive datasets, we aim to reduce biases and accurately enhance the model's ability to recognize emotions across different scenarios and user groups.

Likewise, the real-world application of emotion recognition systems requires careful consideration of ethical implications and privacy concerns. As these systems can potentially extract sensitive emotional states from individuals, there is a need to establish clear guidelines and safeguards to prevent any misuse of this technology. Our proposed framework acknowledges the significance of ethical considerations and promotes the integration of transparency and accountability measures within the model development process. This includes adopting explainable AI techniques to provide insights into how the model arrives at its predictions and allowing users to have control over their data and the inferences drawn from it. By embedding ethical considerations into the core of our approach, we aspire to ensure that responsible and trustworthy deployment practices accompany the benefits of improved sentiment classification from visual data.

Our proposed framework aims to enhance emotion recognition from visual data using a synergistic approach that combines Convolutional Neural Networks (CNNs) and advanced techniques like transfer learning. By addressing current limitations in deep learning methods, we seek to achieve more accurate sentiment classification from images and videos by considering the following objectives:

- **Architectural Fusion for Enhanced Feature Extraction:** Our first objective involves the integration of CNN-based deep learning architectures to extract intricate features from visual data.
- **Mitigating Bias and Ensuring Ethical Deployment:** The second objective focuses on dataset diversity and ethical considerations.

The structure of this paper is meticulously designed to present a cohesive progression of our research endeavor. It begins with elucidating the background setting the stage by highlighting the challenges and opportunities inherent in emotion recognition from visual data. Following this, the paper delves into a comprehensive literature survey that encapsulates existing knowledge related to deep learning techniques, emotion recognition, and sentiment analysis. The subsequent section meticulously outlines the experimental setup, providing details about the chosen visual datasets, the architecture of the employed CNNs, and the incorporation of transfer learning techniques. Finally, the paper culminates with an exhaustive presentation of the results and their subsequent discussion.

II. BACKGROUND

Sentiment analysis has undergone a substantial evolutionary journey, as depicted in Table I. This historical progression spans from the early rule-based systems to the emergence of deep learning models and multimodal analysis techniques. Throughout its development, sentiment analysis has evolved to embrace more sophisticated methodologies, empowering the analysis of emotions and opinions with increasing precision and complexity.

Sentiment analysis plays a crucial role across various industries and for individuals, and its absence would result in severe negative impacts on different aspects of society [5]. In Business and Marketing, understanding customer's perceptions and opinions about products or services through sentiment analysis is indispensable. With it, businesses could leverage customer feedback, leading to a decline in product improvement and effective customer service, ultimately affecting customer satisfaction and revenue [6]. In Politics, grasping public sentiment is pivotal for political parties to understand better their constituents, and government organizations can utilize sentiment analysis to gauge the public's response to policy decisions or changes. In Healthcare, sentiment analysis proves valuable in monitoring and analyzing patient emotions, particularly in mental health and rehabilitation, enabling timely interventions for depression or anxiety [7]. Social Media platforms heavily rely on user engagement and sentiment analysis to analyze feedback, identify trends, and offer personalized content. The absence of sentiment analysis could hinder their ability to

provide tailored recommendations and insights based on user preferences. In the Entertainment industry, sentiment analysis is used to comprehend audience preferences, leading to content customization and improved user experiences [8]. Its absence may hinder the efficiency of content creation and

distribution. Overall, sentiment analysis is essential for informed decision-making in businesses, politics, Healthcare, social media, and entertainment, impacting society and individuals profoundly.

TABLE I. HISTORICAL PERSPECTIVE OF SENTIMENTAL ANALYSIS

Timeline	Approaches	Description
Pre-2000s	Early Sentiment Analysis Techniques	These techniques used simple rule-based systems to generate sentiment scores for texts based on the presence of certain keywords or phrases with positive or negative connotations.
2000s	Machine Learning-Based Approaches	These techniques relied on natural language processing (NLP) and machine learning algorithms to analyze text data. They used supervised learning algorithms such as Naive Bayes, SVMs, and decision trees to classify text into positive, negative, or neutral.
Mid-2000s	Aspect-based Sentiment Analysis	The underlying concept of this approach involves conducting a detailed analysis of text data, delving into a more granular level. It accomplishes this by dissecting the overall sentiment of a text and discerning the sentiment associated with each specific aspect within the text.
2010s	Deep Learning-Based Approaches	Deep neural networks, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), were employed in these approaches to extract features from text data and accurately classify the overall sentiment expressed in the text.
Current	Multimodal and Cross-lingual Sentiment Analysis	These approaches aim to analyze sentiment in multiple languages and through different modalities (text, audio, and gesture recognition). This new approach is built on machine translation and multitask learning architecture to enable sentiment analysis in languages other than English. This has led to the developing of more advanced models that perform complex multimodal analysis across multiple languages.

Sentiment analysis plays a pivotal role in various sectors, empowering businesses, individuals, and industries to make informed decisions that contribute to improving society and its constituents [9]. With sentiment analysis, valuable prospects for improved customer engagement, tailored content delivery, and favorable outcomes on social and economic fronts might be noticed. This absence could hinder progress, resulting in diminished access to personalized experiences and a potentially detrimental impact on societal well-being and economic growth. The integration of sentiment analysis thus emerges as a crucial tool with far-reaching implications, offering a proactive means to harness sentiment insights for the collective benefit.

III. LITERATURE REVIEW

Sentiment analysis is gaining popularity as a prominent research area within natural language processing, attracting numerous studies. Initially, the field relied on rule-based methods to determine the sentiment of texts using specific keywords or phrases [10]. Yet, the effectiveness of these approaches was constrained by their inability to grasp intricate language nuances and variations, prompting the exploration of more advanced techniques. Subsequent research in sentiment analysis witnessed a shift towards machine learning-based approaches, where supervised learning algorithms were utilized to categorize text into positive, negative, and neutral classes [11]. These methods demonstrated improved performance compared to rule-based techniques; however, they still faced limitations in effectively analyzing more intricate linguistic structures.

During the mid-2000s, aspect-based sentiment analysis emerged as a novel approach to assess the sentiment of specific aspects within a text. This method provided a more nuanced and detailed understanding of the sentiments expressed, proving particularly effective in analyzing product reviews [12]. Focusing on individual aspects enabled a comprehensive analysis of various sentiments within a text,

leading to valuable insights and enhanced accuracy in sentiment assessment. In the 2010s, deep learning techniques like Convolutional Neural Networks (CNNs) surfaced and substantially increased sentiment analysis accuracy [13]. These approaches significantly improved the identification of text sentiments and exhibited robustness in handling diverse textual data types. The utilization of deep learning models marked a notable progression in the field, enabling more precise and reliable sentiment analysis results across various text formats.

Sentiment analysis studies have recently expanded to include multimodal and cross-lingual analysis, moving beyond traditional text-based methods. These advanced approaches leverage machine learning to analyze emotions conveyed through various multimedia forms, such as audio, video, and images [14]. This evolution from rule-based and machine-learning-based approaches to deep learning techniques has significantly improved sentiment analysis accuracy and efficiency, benefiting industries like healthcare, advertising, and entertainment [15]. Further research in this field holds great potential for developing even more sophisticated models, enhancing sentiment analysis effectiveness across diverse applications.

IV. PROPOSED METHODOLOGY

To effectively classify sentiments from visual information, developing a robust learning architecture model is essential as show in Fig. 2. This necessitates a thorough understanding of the diverse features and cues that can convey emotions in visual content, including facial expressions, body language, and color schemes. One potential approach to constructing a learning architecture model for sentiment analysis involves employing a deep neural network. Such a network can be trained on large datasets containing labeled visual content, enabling it to recognize patterns and correlations between specific features and emotions. This model type can be fine-tuned to accommodate various types of visual content, such as

images, videos, or live streams, tailored to the specific application.

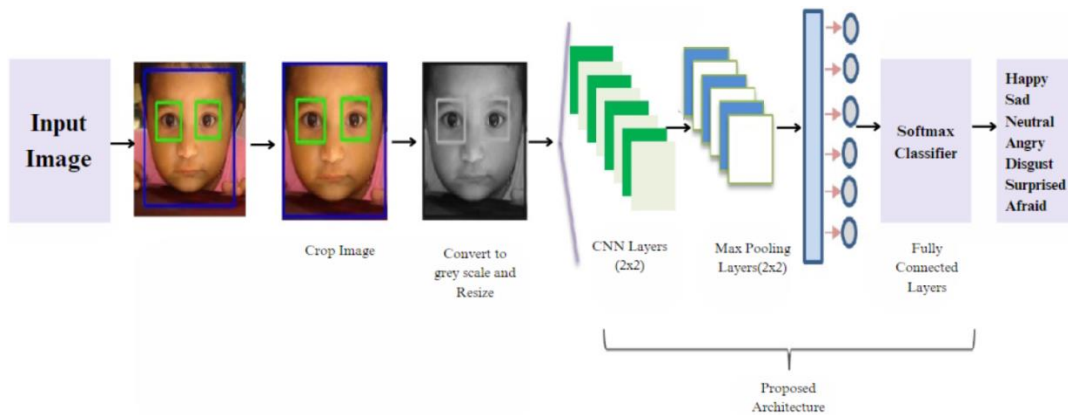


Fig. 2. Deep learning architecture for advanced sentiment analysis.

A. Importing Libraries

In our research on sentiment detection based on vision analysis, we begin by importing essential libraries that will empower us to build and evaluate our models effectively. Tensor Flow and Keras form the backbone of our deep learning infrastructure, enabling us to construct and train complex neural networks for sentiment analysis. NumPy is indispensable for numerical computations and data manipulation, ensuring efficient handling of image data and feature extraction. Open CV plays a pivotal role in image processing tasks, aiding us in pre-processing visual data and extracting meaningful features like facial expressions and color information. Lastly, Matplotlib is instrumental in visualizing and presenting our results in a clear and informative manner, facilitating better insights into the performance of our sentiment analysis models. Together, these libraries form the foundation of our research, enabling us to explore and implement cutting-edge techniques for sentiment detection through vision analysis.

B. Displaying Sample Images

In the context of sentiment detection using vision analysis, displaying a selection of sample images from the dataset is crucial to gain insights into the visual content and the emotions conveyed in the images. By examining these sample images, researchers can better understand the diversity and complexity of the data they will be working with. It allows them to identify different facial expressions, body language, color schemes, and other visual cues that play a role in conveying sentiments. Additionally, this step aids in identifying potential challenges, such as variations in image quality, lighting conditions, and the representation of different emotions, which can impact the accuracy of the sentiment detection model. By visually inspecting the sample images, researchers can ensure that the dataset is diverse, representative, and suitable for effectively training and evaluating their sentiment detection system. This process helps researchers make informed decisions about data preprocessing, feature extraction, and model development, ultimately contributing to the success and reliability of the sentiment detection project.

C. Training and Validation Data

To ensure the efficacy and generalization of our sentiment detection model in the context of vision analysis, we divide our dataset into two essential subsets: the training set and the validation set. The training set serves as the foundation for training our model, allowing it to learn from various images with various sentiments. This step is critical for the model to comprehend and recognize the intricacies of different emotions expressed in visual content. The validation set, on the other hand, is employed to assess the performance of the trained model. By evaluating the model's accuracy and efficiency on the validation set, we can validate its ability to detect sentiments in new and unseen visual data. Ensuring that both sets have a representative distribution of images with different sentiments is crucial to prevent biased training and to enhance the model's robustness. This careful division of data allows us to create a well-performing sentiment detection system that can effectively handle a wide array of visual content, contributing to a successful sentiment analysis based on vision analysis.

D. Model Building

The model is constructed using a sequential algorithm, which entails a linear stack of layers. The architecture comprises four Convolutional Neural Network (CNN) layers, allowing the model to learn hierarchical features from the visual data. Following the CNN layers, two fully connected layers enable the model to comprehend complex relationships between the learned features and sentiment classes. To obtain probability distributions over the different sentiment categories, the SoftMax activation function is used in the last layer, ensuring that the output represents the likelihood of each sentiment class. The ADAM optimizer is employed for model optimization, as it effectively adapts the learning rate and aids in achieving faster convergence during the training process. This carefully designed model architecture leverages the power of CNNs in feature extraction from visual data, culminating in a sentiment detection system capable of accurately recognizing and classifying emotions expressed in images or videos.

E. Fitting the Model with Training and Validation Data

To effectively build and optimize the sentiment detection model in our vision analysis project, we train the model using the training data and validate its performance using the validation data. This step involves adjusting crucial hyperparameters, such as batch size, learning rate, and the number of epochs, to fine-tune the model's performance. The batch size determines the number of training examples used in each iteration, while the learning rate governs the step size during model optimization, impacting the convergence speed and overall performance. The number of epochs defines the number of times the model iterates through the training data. By carefully adjusting these hyperparameters, we aim to balance model underfitting and overfitting, optimizing the model's ability to generalize to new and unseen data. This iterative process enables us to find the best configuration that maximizes the model's accuracy and efficiency, ultimately leading to a robust and reliable sentiment detection system for visual content analysis.

F. Calculating Training Loss and Validation Loss

As we proceed with training the sentiment detection model in our vision analysis project, it is essential to closely monitor the training and validation loss throughout the training process. The training loss represents the error between the model's predictions and the actual sentiment labels on the training data, while the validation loss measures the model's performance on unseen data from the validation set. Plotting the loss curves enables us to visualize the convergence of the model and detect potential issues of overfitting or underfitting. An optimal model should exhibit a decrease in both training and validation loss, indicating that it is learning to generalize well to new data. However, if the training loss continues to decrease while the validation loss starts to increase or plateaus, it could be a sign of overfitting, where the model memorizes the training data rather than learning general patterns. Conversely, if both the training and validation losses remain high, it may indicate underfitting, suggesting that the model needs to capture the underlying complexities of the data. By analyzing the loss curves, we can make informed decisions on adjusting the model architecture or hyperparameters to strike the right balance and achieve a well-performing sentiment detection system capable of accurately analyzing emotions in visual content. This iterative process ensures that the model is trained effectively and is robust enough to handle diverse data, enhancing the project's overall success.

G. Export the Model

Upon successful training and evaluating the sentiment detection model in our vision analysis project, it is crucial to export the model for future use and deployment. This involves saving the model's architecture and the learned weights in a file format compatible with our framework, such as HDF5 or Saved Model. By doing so, we preserve the entire model configuration and the knowledge acquired during training, allowing us to reuse the model for sentiment analysis on new, unseen visual data. Exporting the model in a compatible format ensures easy integration into different applications or platforms, enabling seamless utilization in real-world

scenarios. Moreover, this step facilitates collaboration with other researchers or teams using the exported model to perform sentiment analysis on their specific datasets or tasks. In essence, exporting the model is a critical step in turning our research efforts into a practical and valuable tool that can be readily applied in various domains requiring sentiment analysis from visual content.

H. Real-Time Sentiment Detection using OpenCV

Incorporating Real-Time sentiment detection into our vision analysis project involves implementing the exported model with the OpenCV library. By leveraging OpenCV's capabilities, we can seamlessly capture live video streams from a webcam or video source. The exported model, comprising the architecture and learned weights, is then utilized to analyze the emotions expressed in the Real-Time visual content. As each video frame is processed through the model, sentiments are rapidly detected and classified. This enables the system to provide immediate feedback on the emotional content displayed in the video feed, offering valuable insights into the sentiments expressed by individuals or subjects. This Real-Time sentiment detection empowers us to understand and respond to emotional cues in live scenarios, making it applicable in various applications, such as Real-Time audience feedback analysis, user experience evaluation, and emotion-aware interactive systems. By merging the exported model with OpenCV, we create an efficient and powerful tool that can continuously and accurately perform sentiment analysis in Real-Time video streams, bringing practicality and real-world value to our vision analysis research.

V. RESULTS AND DISCUSSION

In our sentiment analysis study using deep learning models, we presented the outcomes and insights achieved through our approach, which enabled the recognition of emotions such as sadness, happiness, neutrality, and fear in visual data as show in Fig. 3. Our sophisticated deep-learning architecture leveraged Convolutional Neural Networks (CNNs) to extract meaningful features, allowing for accurate sentiment classification.

The performance of our deep learning model surpassed our expectations, achieving a notable accuracy of 84.6% (Table II) in detecting various emotions expressed in visual content. We are optimistic that this accuracy can be further improved based on specific system requirements. By training the model with more images and increasing the number of CNN and Pooling layers, we can enhance its ability to generalize across different emotional expressions and boost its accuracy.

TABLE II. COMPARISON OF MULTIMODAL DEEP LEARNING APPROACH FOR REAL-TIME SENTIMENT ANALYSIS IN VIDEO STREAMING WITH OTHER WORKS

SI No	Year	Method Used	Accuracy	Reference
1	2017	CNN-RNN Ensemble for Videos	72.5%	[16]
2	2018	Fusion of Audio-Visual Features	67.8%	[17]
3	2020	Multimodal DL for Video Streaming	75.2%	[18]
4	2023	CNN Multimodal for	84.6%	Our Work

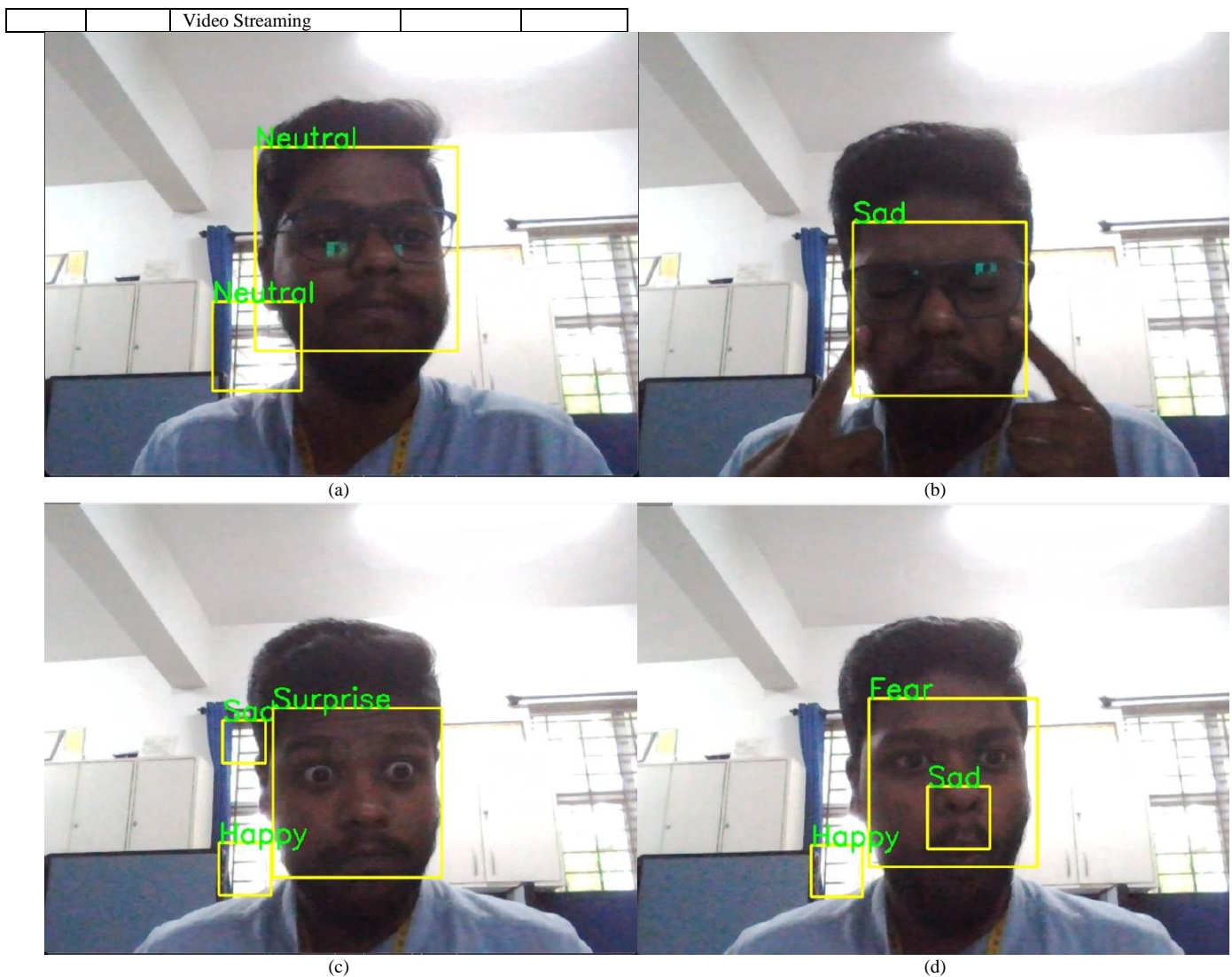


Fig. 3. Sentimental recognition a) Neutral, b) Sad, c) Surprise, d) Fear.

We also incorporated multimodal and cross-lingual analysis in our study, enhancing the versatility of our sentiment analysis system. By including multiple modalities like audio, video, and image data, we gained a comprehensive understanding of sentiments conveyed in diverse forms of multimedia. This multimodal analysis provided richer insights into the emotional context, enabling a deeper exploration of cross-lingual sentiments and expressions beyond traditional text-based approaches.

The development of our deep learning-based sentiment analysis model signifies a significant advancement in the field. The model's promising accuracy and efficiency hold great potential for real-world applications, particularly in the healthcare, advertising, and entertainment industries. By accurately detecting and interpreting emotions expressed in visual data, our model opens up new possibilities for understanding user preferences, improving customer experiences, and enhancing content personalization.

VI. CONCLUSION

Our research on sentiment analysis using the novel multi-view deep learning framework has demonstrated remarkable success in recognizing emotions from visual data. The deep learning model achieved an impressive accuracy of 84.6% in accurately detecting various emotions, including sadness, happiness, neutrality, and fear. Further refinement, such as increasing the number of CNN and pooling layers and incorporating more extensive training datasets, could lead to even higher accuracy levels. The integration of multimodal and cross-lingual analysis in our study has enriched the versatility of our sentiment analysis system, providing valuable insights into sentiments expressed across diverse forms of multimedia. With its promising performance, our deep learning-based sentiment analysis model holds substantial promise for practical applications in the healthcare, advertising, and entertainment industries. Its ability to discern emotions from visual content opens up new avenues for improving customer experiences, enhancing content personalization, and gaining a deeper understanding of user

preferences. We foresee the widespread application of our framework in various industries, benefiting user engagement, customer service, and content personalization. While this research marks a significant stride in sentiment analysis, we acknowledge the scope for further improvement as we remain committed to advancing the field and contributing to a nuanced understanding of human emotions in the digital era.

REFERENCES

- [1] Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*, 83, 19-52.
- [2] Kardakis, S., Perikos, I., Grivokostopoulou, F., & Hatzilygeroudis, I. (2021). Examining attention mechanisms in deep learning models for sentiment analysis. *Applied Sciences*, 11(9), 3883.
- [3] Singh, S. K., Thakur, R. K., Kumar, S., & Anand, R. (2022, March). Deep learning and machine learning based facial emotion detection using CNN. In *2022 9th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 530-535). IEEE.
- [4] Tolan, S., Pesole, A., Mart'inez-Plumed, F., Fern'andez-Mac'ias, E., Hern'andez-Orallo, J., & G'omez, E. (2021). Measuring the occupational impact of ai: tasks, cognitive abilities and ai benchmarks. *Journal of Artificial Intelligence Research*, 71, 191-236.
- [5] Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, 145, 103724.
- [6] Caviggioli, F., Lamberti, L., Landoni, P., & Meola, P. (2020). Technology adoption news and corporate reputation: Sentiment analysis about the introduction of Bitcoin. *Journal of Product & Brand Management*, 29(7), 877-897.
- [7] Yan, C., Liu, J., Liu, W., & Liu, X. (2022). Research on public opinion sentiment classification based on attention parallel dual-channel deep learning hybrid model. *Engineering Applications of Artificial Intelligence*, 116, 105448.
- [8] Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3, 100073.
- [9] Nakayama, M., & Wan, Y. (2019). The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Information & Management*, 56(2), 271-279.
- [10] Goularas, D., & Kamis, S. (2019, August). Evaluation of deep learning techniques in sentiment analysis from twitter data. In *2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)* (pp. 12-17). IEEE.
- [11] Shamrat, F. M. J. M., Chakraborty, S., Imran, M. M., Muna, J. N., Billah, M. M., Das, P., & Rahman, O. M. (2021). Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*, 23(1), 463-470.
- [12] Buzova, D., Sanz-Blas, S., & Cervera-Taulet, A. (2019). Does culture affect sentiments expressed in cruise tours' eWOM?. *The Service Industries Journal*, 39(2), 154-173.
- [13] Puzyrev, V. (2019). Deep learning electromagnetic inversion with convolutional neural networks. *Geophysical Journal International*, 218(2), 817-832.
- [14] Uymaz, H. A., & Metin, S. K. (2022). Vector based sentiment and emotion analysis from text: A survey. *Engineering Applications of Artificial Intelligence*, 113, 104922.
- [15] Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., ... & Ashraf, I. (2022). Sentiment analysis on Twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems*, 255, 109780.
- [16] Farhoudi, Z., & Setayeshi, S. (2021). Fusion of deep learning features with mixture of brain emotional learning for audio-visual emotion recognition. *Speech Communication*, 127, 92-103.
- [17] Tsai, Y. H. H., Bai, S., Liang, P. P., Kolter, J. Z., Morency, L. P., & Salakhutdinov, R. (2019, July). Multimodal transformer for unaligned multimodal language sequences. In *Proceedings of the conference. Association for Computational Linguistics. Meeting (Vol. 2019, p. 6558)*. NIH Public Access.
- [18] Zellers, R., Lu, X., Hessel, J., Yu, Y., Park, J. S., Cao, J., ... & Choi, Y. (2021). Merlot: Multimodal neural script knowledge models. *Advances in Neural Information Processing Systems*, 34, 23634-23651.

3D Magnetic Resonance Image Denoising using Wasserstein Generative Adversarial Network with Residual Encoder-Decoders and Variant Loss Functions

Hanaa A. Sayed¹, Anoud A. Mahmoud^{2*}, Sara S. Mohamed³

Dept. of Computer Science-College of Computer Science and Engineering, Taibah University, KSA-Dept. of Mathematics, Faculty of Computers and Information, Assiut University, 71516 Assiut, Egypt¹
Dept. of Mathematics and Computer Science-Faculty of Science, New Valley University, 72511 New Valley, Egypt^{2,3}

Abstract—Magnetic resonance imaging (MRI) is frequently contaminated by noise during scanning and transmission of images, this deteriorates the accuracy of quantitative measures from the data and limits disease diagnosis by doctors or a computerized system. It is common for MRI to suffer from noise commonly referred to as Rician noise because the uncorrelated Gaussian noise is present in both the real and imaginary parts of a complex K-space image with zero mean and equal standard deviation, the distribution of noise in magnitude MR images typically tends to be related to Rician distributions. To remove the Rician noise from an MRI scan, deep learning has been used in the MRI denoising method to achieve improved performance. The proposed models were inspired by the Residual Encoder-Decoder Wasserstein Generative Adversarial Network (RED-WGAN). Specifically, the generator network is residual autoencoders combined with the convolution and deconvolution operations, and the discriminator network is convolutional layers. As a result of replacing Mean Square Error (MSE) in RED-WGAN with Structurally Sensitive Loss (SSL), RED-WGAN-SSL has been proposed to overcome the loss of important structural details that occurs because of over-smoothing the edges. The RED-WGAN-SSIM model has also been developed using Structural Similarity Loss SSIM. The proposed RED-WGAN-SSL and RED-WGAN-SSIM models are formed by using the SSL, SSIM, Visual Geometry Group (VGG), and adversarial loss that are incorporated to form the new loss function. They preserved the informative details and fine image better than RED-WGAN, so our models could effectively reduce noise and suppress artifacts.

Keywords—Deep learning; image denoising; MRI; Wasserstein GAN; loss function

I. INTRODUCTION

MRI is a medical imaging process that produces multidimensional images of the inside of the body; it uses powerful magnets and radio waves generated by computers rather than injecting contrast agents. It is considered one of the most attractive modalities that have been used in the diagnosis and treatment of several neurological diseases because it can show 3D details of internal living tissues and the human body organs. MRI plays an increasingly important role in pathological and physiological diagnostics and scientific

research. Physiological noise impedes the acquisition of signals and contaminates raw data sets by artificial outliers. As a result of this practical issue, more advanced technologies have difficulty being applied in clinical research. Increasing noise levels may have a bad effect not only on the accuracy of computed diagnostic systems, but also on manual disease inspection and the reliability of quantitative image processing including segmentation, registration, visualization, super-resolution, and classification [1]. Raw data is usually polluted by White Additive Gaussian Noise (WAGN) in the real and imaginary parts. This noise is assumed to have equal variance and zero mean in the entire K space of the data, meaning that it affects both the real and imaginary parts of the data equally. Rician noise, on the other hand, is signal-dependent, which makes it harder to separate from the signal and can result in biased estimates of image intensity. Additionally, in high SNR regions, the Rician noise is close to the Gaussian distribution. To achieve reliable analysis results, it is necessary to remove noise before performing further image processing.

In MRI denoising, the goal is to effectively restore a clean image from a contaminated MR image and preserve valuable information [2]. In the past, many research attempts for MRI denoising were made to remove additive noise, most of which used the Rician noise model. In general, these methods can be categorized into three types: spatial filtering, transform domain filtering, and statistical methods [3]. The spatial domain techniques are directly applied to image pixels [4]. There are several traditional spatial image filters, including median [5], Gaussian [6], Wiener [7], diffusion [8], and bilateral filters. Anisotropic diffusion filter [9] significantly retained informative details of edges and reduced the noise from the images by smoothing local regions in the image, but the image was still blurry. This filter tried to avoid blurring of the edges by utilizing the edge-stopping function. A transform domain image filter is different from a spatial domain image filter. In that transform domain filtering methods first transform the space domain into another domain, and then they process the transformed image in the new domain based on the different characteristics of the image and its noise such as the frequency and wavelet domains [10]. Rician noise in MRI data has been successfully denoised by well-known block matching 3D

(BM3D) [11]. Higher-order singular value decomposition (HOSVD) [12] was developed to denoise MR volume data, and its performance was improved compared to BM3D.

Deep Learning (DL) made impressive progress in image processing and computer vision fields by introducing new effective methodologies. It has been used on low-level tasks to denoise [13,14], deblur [15], and restore super-resolution images [16]. CNNs and autoencoders achieved competitive performance with state-of-the-art methods, such as BM3D and NLM, for image denoising [14]. Lore et al. [17] developed LLNet, a deep auto-encoder that enhances contrast and removes noise. Zhang et al. [14] used Denoising convolutional neural networks (DnCNNs) to handle Gaussian denoising with unknown noise levels, which is different from traditional discriminative models that are trained specifically for certain noise levels. DnCNNs not only achieved excellent performance quantitatively and qualitatively by using residual learning strategy but also to speed up the training process on GPU computing by using batch normalization (BN) [18]. Zhang et al. [19] proposed a new fast, flexible CNN denoising model namely FFDNet. FFDNet can handle a wide range of noises, remove white Gaussian and spatially variant noise which requires a noise level map, and is faster than BM3D. It is effective and provides a practical solution to denoising applications because it achieves a good balance between performance and inference speed.

Although researchers have made great efforts in MRI denoising to retrieve free noise images and get effective results, the research on MRI denoising is quite limited. Current methods suffer from several drawbacks including nonlinear optimization, tuning the parameters of neural networks, high computations, and/or sensitive parameters, which seriously lead to unsatisfactory denoising results. In this work, to avoid these problems, the proposed models are inspired by an MRI denoising method based on RED-WGAN [3]. This paper mainly contributes to learning the distribution of data in a low-dimensional manifold using the WGAN framework, different loss functions such as VGG loss [20], SSL loss, SSIM loss [21], residual networks and autoencoders [22], which were employed in the proposed models to preserve clinical relevant details such as the edges and the informative structure. MSE loss in the RED-WGAN model has been replaced with SSL loss to overcome the loss of important structural details occurring due to over-smoothing edges. Also, SSIM loss has been used to preserve the image details in high resolution. The proposed method is computationally fast and can be implemented on Graphic Processing Units (GPUs). The rest of this paper is organized as follows: Section II defines the related work; Section III presents the proposed models; Section IV describes the experiments and results; and finally, Section V shows the conclusions and future work.

II. RELATED WORK

In the field of clinical imaging, Jiang et al. [23] proposed a multichannel convolutional neural network (MCDnCNN) for MRI denoising with and without a specific noise level, in which CNN layers were combined with residual learning [24] and VGG network architecture. It robustly denoises 3D MR images with Rician noise. Manjon et al. [25] proposed a two-

stage approach to effectively reduce the noise: the non-local PCA thresholding strategy is used to filter the noisy image by automatically estimating the local noise level in the image; then this filtered image is used as a guide in the rotationally invariant NLM [prefiltered rotationally invariant nonlocal means (PRI-NLM)] filter. Ran et al. [3] introduced the RED-WGAN model for MRI denoising, which consists of three main parts: the generator network, the discriminator network, and combined loss functions. In the generator network, the residual autoencoder structure is composed of convolutional and deconvolutional layers symmetrically. The discriminator network consists of convolutional layers. The authors combined three loss functions including the MSE loss function [21], Adversarial loss, and VGG loss. The proposed model powerfully reduced the noise and retrieved the structural details. Tripathi et al. [26] proposed a novel CNN-DMRI model to remove the Rician noise from MRI, which utilized a set of convolutional layers to capture the image features while the noise is separating. As part of CNN-DMRI structures, encoder-decoder structures were also employed to retain the informative features of the image while unnecessary ones are ignored. The qualitative and quantitative results of the proposed method are promising. Li et al. [1] successfully applied Rician denoising with a progressive learning approach to MR images. The progressive network, called RicianNet, consists of two sub-RicianNets, which are residual blocks: one of the sub-networks fitted the noise distribution at the pixel-domain without batch normalization layer, and the other one employs ResNet structure with batch normalization layer in the feature domain, thus enhancing the nonlinear mapping. The authors improved the network performance by employing the BN layer, Convolutional layer, and residual unit. RicianNet had better quantitative measures and significant improvements in visual inspections. Aetesam et al. [27] proposed a deep neural architecture for MRI denoising to remove Gaussian-impulse noise by using an ensemble-based residual learning strategy. The proposed model achieved high-quality visual results and high quantitative metrics compared to other state-of-the-art models. Gregory et al. [28] developed a multi-branch deep neural network architecture, HydraNet, to remove noise from MR images at a wide range of noise levels. Compared to other deep learning-based methods, the HydraNet network demonstrated powerful results in the denoising of complex noise distributions. Wu et al. [2] used 3D Parallel-RicianNet for 3D MRI denoising, which combines global and local information for noise reduction. To expand its receptive field, the authors introduced a powerful module called dilated convolution residuals (DCR).

III. PROPOSED DENOISING MODEL

It is difficult to denoise an MRI because magnitude images, which consist of real and imaginary parts, are commonly used [3, 29]. The noise in the magnitude MR image follows a Rician noise distribution [29], which is significantly more complicated than traditional additive Gaussian noise.

In MRI denoising, a free MR image is obtained by removing noise from a noisy MR image, as follows:

$$z = \theta(x) \quad (1)$$

Where z denotes a noisy MR image, x denotes the corresponding noise-free MR image and $(x, z \in R^{H \times W \times D})$, and function θ maps to the noise. The model-based DL is independent of noise and its statistical characteristics since it is a black box. So, the denoising process aims to approximate the function θ^{-1} to the possible optimal and can be expressed as follows:

$$\operatorname{argmin}_f \|\hat{z} - z\|_2^2 \quad (2)$$

where $\hat{z} = Q(z)$, which corresponds to an estimate of x , and Q indicates the optimal approximation of θ^{-1} [3].

A. Wasserstein GAN

The GANs model can be described in Eq. (3).

$$\min_G \max_D V(D, G) = E_{x \sim P_r(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3)$$

The two variables x and z can be interpreted as samples drawn from two different distributions of data, which in the context of statistics, can be considered as being real image distributions p_r and noisy image distributions p_z respectively. Then, the denoising function moves the samples from p_z to p_g which is close to p_r . An Adversarial Generative Network (GAN) is made up of two networks, a generator, G , and a discriminator, D . There have been many uses of GAN in research fields such as computer vision [30,31], security [32], and data generation [33]. The generator generates samples from random noise as close as possible to real data to deceive a discriminator. The discriminator attempts to distinguish between the two-distribution of the generative model p_g and the real data p_r .

Despite its success in image generation, GAN suffers from training instabilities, extremely sensitive parameter tuning, vanishing gradient, and mode collapse [34]. It has been proposed to improve GAN by using Wasserstein GAN (WGAN) [35]. The loss function of WGAN was proposed to avoid vanishing gradients. Wasserstein Distance measures the divergence between real distribution P_r and model distribution P_g ; In WGAN, weight clipping is used to enforce Lipschitz constraints, when clipping parameters are too small or too large can result in the same original GAN problems. Therefore, the Gradient penalty (GP) was used instead of weight clipping to enforce the Lipschitz constraint on the critic(discriminator) during training.

WGAN-GP is a WGAN with a gradient penalty, and the loss function is shown in Eq. (4).

$$L = E_{\tilde{x} \sim p_g} [D(\tilde{x})] - E_{x \sim p_r} [D(x)] + \lambda E_{\hat{x} \sim p_{\hat{x}}} [(\|\nabla_{\tilde{x}} D(\hat{x})\|_2 - 1)^2] \quad (4)$$

The sample of $p_{\hat{x}}$ is taken uniformly between two points sampled from P_r and P_g , the last term is a gradient penalty factor, and λ is a penalty coefficient.

B. Loss Functions

Mean Squared Error (MSE) loss or $L2$ calculates the normalized Euclidean distance between a generated patch $G(z)$ from model distribution P_g and the patch of noise-free images x from real distribution P_r ; it minimizes the pixel-wise difference

between them [21]. Recent studies suggest that although the per-pixel MSE results have a high peak signal-to-noise ratio (PSNR), it may cause the loss of some important structural details due to an over-smoothed edge. The formula of $L2$ loss is expressed as in Eq. (5).

$$L2 = \frac{1}{HWD} \|G(z) - x\|_2^2 \quad (5)$$

The Perceptual Loss (PL) was used to overcome this problem by being employed in the feature space instead of directly estimating MSE on a pixel-by-pixel basis. A pre-trained VGG-19 network [20] can be applied to extract the features from the generated patch and noise-free patch, VGG loss compares high-level perceptual differences [21].

$$PL(G) = E_{(x,z)} \frac{1}{HWD} \|\varphi G(z) - \varphi(x)\|_2^2 \quad (6)$$

In which φ is a feature extractor, W refers to the width, H indicates to the height, and D is the depth of feature maps. The perceptual loss can be described as the following formula:

$$L_{VGG}(G) = PL(G) E_{(x,z)} \frac{1}{HWD} \|VGG(G(z)) - VGG(x)\|_2^2 \quad (7)$$

Structural Similarity (SSIM) loss measures the similarity between two patches $G(z)$ and x based on three comparisons: contrast, luminance, and structure [21]. The SSIM can perform better than the MSE in perceptual pattern recognition because it is visually based. The original SSIM is formulated as follows in Eq. (8).

$$SSIM(x, y) = \frac{2\mu_x\mu_y + C1}{\mu_x^2 + \mu_y^2 + C1} * \frac{2\sigma_{xy} + C2}{\sigma_x^2 + \sigma_y^2 + C2} = L(x, y) * cs(x, y) \quad (8)$$

Where μ_x , μ_y , σ_x , σ_y and σ_{xy} are the means, standard deviations, and the cross-covariance of the two images (y, x) obtained from the model and the corresponding noise-free image respectively and $C1$, $C2$ are constants [21,36,37]. If x and y are very similar, SSIM approaches 1.

$$L_{SSIM} = 1 - SSIM(x, y) \quad (9)$$

In this paper, we presented RED-WGAN-SSL and RED-WGAN-SSIM models based on WGAN. They are incorporated with different loss functions to reduce the noise in 3D MRI and retain structural information. The two proposed models are compared with RED-WGAN [3]. The joint loss functions for all models are formulated as follows:

$$L_{RED-WGAN} = \lambda_1 L_{WGAN}(G) + \lambda_2 L_{VGG} + \lambda_3 L_{MSE} \quad (10)$$

$$L_{RED-WGAN-SSIM} = \lambda_1 L_{WGAN}(G) + \lambda_2 L_{VGG} + \lambda_3 L_{SSIM} \quad (11)$$

$$L_{RED-WGAN-SSL} = \lambda_1 L_{WGAN}(G) + \lambda_2 L_{VGG} + \lambda_3 L_{SSL} \quad (12)$$

C. Network Architectures

The proposed models' architecture is inspired by the RED-WGAN architecture [3], which is made up of a G network, a D network, and a VGG network. The G network structure is an autoencoder network that consists of the convolution and deconvolution layers that are symmetrical to deal with the noise. The convolution and deconvolution layer pairs are linked by short connections. The deconvolution layers and the short connections are proposed to speed up the training procedure and maintain more details. There are 8 layers in the

encoder-decoder generator: four convolutional layers and four deconvolutional layers. A 3D convolution is applied to the first seven layers, followed by a batch-normalization and a LeakyReLU, except the last layer, which has a 3D convolution and a LeakyReLU without a batch-normalization; each layer uses $3 \times 3 \times 3$ kernels, the generator employed 32, 64, 128, 256, 128, 64, 32, 1 filter. The VGG network is used to extract features.

The structure of the discriminator network D consists of 3 convolutional layers. All layers perform 3D convolutional operations in sequence with 32, 64 and 128 filters and have $3 \times 3 \times 3$ kernel size, followed by a fully connected layer in the last layer that has a single output.

IV. EXPERIMENTS AND RESULTS

The two proposed models RED-WGAN-SSL and RED-WGAN-SSIM were extensively tested on clinical datasets to validate their performance.

A. Clinical Data

Clinical experiments were conducted using the IXI dataset [38] gathered from three hospitals: Hammersmith Hospital, Guy's Hospital, and the Institute of Psychiatry. The above-mentioned website provides detailed information on scanning configuration. The Hammersmith dataset is a subset of the IXI dataset obtained from a Philips 3T scanner. 110 PD-weighted brain image volumes were randomly chosen. The training set consists of 100 image volumes from the Hammersmith dataset, and the testing set consists of 5 image volumes from the Hammersmith dataset, it also included 5 image volumes from the Guy's Hospital dataset to evaluate the robustness of the proposed models. We manually added Rician noise to the training set and testing set to simulate noisy images. Many training samples are required for deep learning-based methods, which is especially challenging in clinics.

B. Parameter Setting

The training was performed on PD-weighted brain image volumes with specific levels of noise. According to the suggestions in [31,39], the parameters λ_1 , λ_2 , and λ_3 were experimentally set to 1, 0.1, and $1e-3$, respectively. A penalty coefficient λ in Eq. (4) was specified in following the suggestion [35] to 10. The loss function was optimized by the Adam algorithm [40], and the parameters α , β_1 , and β_2 for the optimizer were set to $1e-4$, 0.5, and 0.9, respectively.

C. Results

To evaluate the performance of the proposed denoising models RED-WGAN-SSL and RED-WGAN-SSIM in comparison to RED-WGAN, three quantitative metrics were utilized. The first metric, PSNR, involved comparing the denoised images to the original (ground truth) images by calculating RMSE. The second metric, RMSE, measured the difference between the denoised and ground truth images, lower values indicating better image quality. Lastly, the SSIM was used to compare the similarities between the denoised and ground truth images, taking into account the luminance, contrast, and structure of the images.

1) *Results obtained using a mini-batch size of 11*: This section illustrates the different results for RED-WGAN-SSL, RED-WGAN-SSIM, and RED-WGAN that were trained on PDw images with different levels of noise (5%, 9%, 11%, and 15%). Then, the three denoising models were tested on the same levels of noise (5%, 9%, 11%, and 15%).

a) *Quantitative Results*: Table I presents the average quantitative analysis. The results demonstrate that when the noise level is less than 11%, RED-WGAN-SSL and RED-WGAN-SSIM exhibit slightly better performance than RED-WGAN. As the noise level increases, the performance of RED-WGAN-SSL is mildly better than that of RED-WGAN and RED-WGAN-SSIM.

b) *Qualitative Results*: This section illustrates the different qualitative results for the denoising models RED-WGAN-SSL, RED-WGAN-SSIM and RED-WGAN. Fig. 1 shows results obtained for the PDw brain images in the testing set with 15% Rician noise as the models were also trained on images with 15% Rician noise. Each model suppresses noise to a different degree. However, some vital details are distorted as in RED-WGAN-SSIM. In Fig. 2, it is important to mention that all the models at the noise level of 11% can remove noise to a different degree and that the RED-WGAN-SSL and RED-WGAN-SSIM have better results compared with RED-WGAN, as they preserve more structural details than RED-WGAN as shown by the red arrow. RED-WGAN-SSL suppresses noise better than other models. The results show that the lower the noise level, the better the results and closer to the original reference image as observed at level noise of 9% and 5% in Fig. 3 and Fig. 4 respectively. Consequently, the structure details were preserved while noise was effectively reduced especially at level noise of 5%.

2) *Results obtained using a mini-batch size of 80*: Based on the results obtained in Table II, the RED-WGAN-SSL seems to have performed better in terms of PSNR, SSIM, and RMSE than all the models considered. Fig. 5 provides a visual representation of the different results for RED-WGAN-SSL, RED-WGAN-SSIM, and RED-WGAN on the PDw brain images that were corrupted by 15% Rician noise in the training set and then were tested with 15% Rician noise. It is important to note that all of the models are capable of suppressing noise in converging degrees. The RED-WGAN-SSL model has an improvement in noise suppression compared to the RED-WGAN model as shown in the red arrow, and it also produces results that are more consistent compared to the original reference images. RED-WGAN-SSL analysis results show that most of the noise has been reduced efficiently and the structure details have been retained much better than other models. The quantitative results of different models for Fig. 5 are presented in Table II. There was an agreement between the visual inspection and quantitative results in terms of PSNR, SSIM, and RMSE when using RED-WGAN-SSL, which is the best result of all the modalities.

TABLE I. A COMPARISON OF PSNR, SSIM, AND RMSE METRICS ON DENOISED PDW EXAMPLE WITH DIFFERENT LEVELS OF RICIAN NOISE FROM THE TESTING SET

		5%	9%	11%	15%
Noise	PSNR	41.853721	46.758666	44.828463	41.853721
	SSIM	0.306874	0.192677	0.157860	0.109305
	RMSE	0.374527	0.678883	0.831107	1.135392
RED-WGAN	PSNR	63.428257	59.420627	57.281444	54.901224
	SSIM	0.802056	0.760506	0.699665	0.605663
	RMSE	0.116628	0.180583	0.229623	0.303276
RED-WGAN-SSL	PSNR	63.783810	59.621328	57.85996	55.001128
	SSIM	0.816764	0.761466	0.721758	0.588793
	RMSE	0.112812	0.177403	0.215451	0.301685
RED-WGAN-SSIM	PSNR	63.617620	59.364931	57.574540	54.617207
	SSIM	0.787149	0.726462	0.718152	0.560422
	RMSE	0.114135	0.181923	0.222884	0.313424

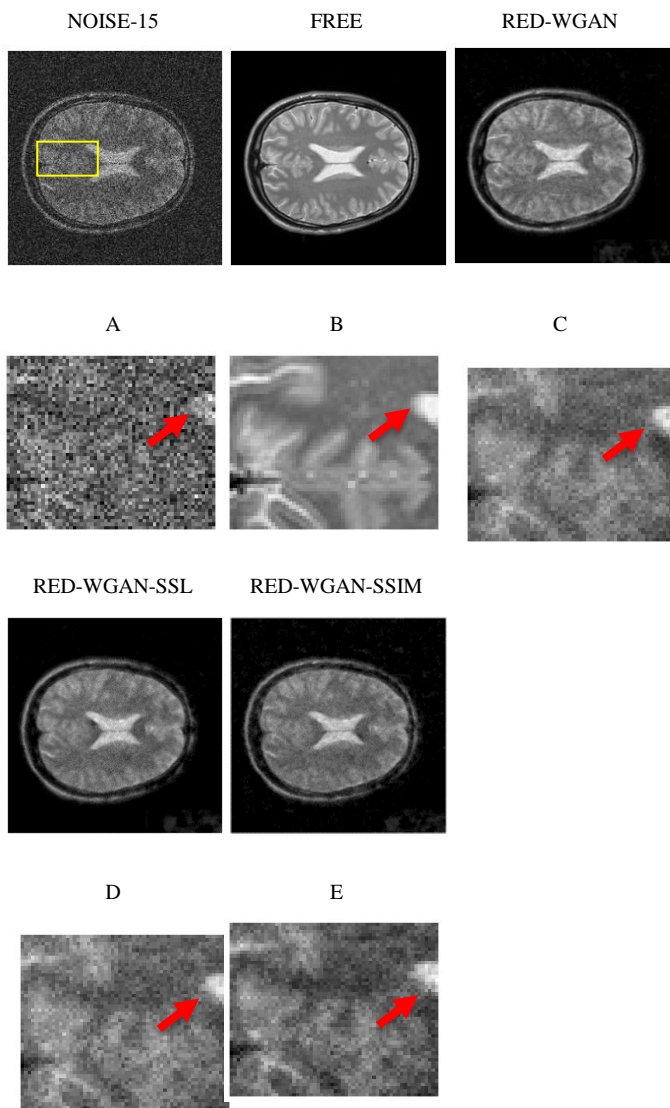


Fig. 1. Denoised PDw example with 15% Rician noise from the testing set at a mini-batch size =110: (A) Noisy image, (B) Ground truth image, (C) RED-WGAN, (D) RED-WGAN-SSL and (E) RED-WGAN-SSIM.

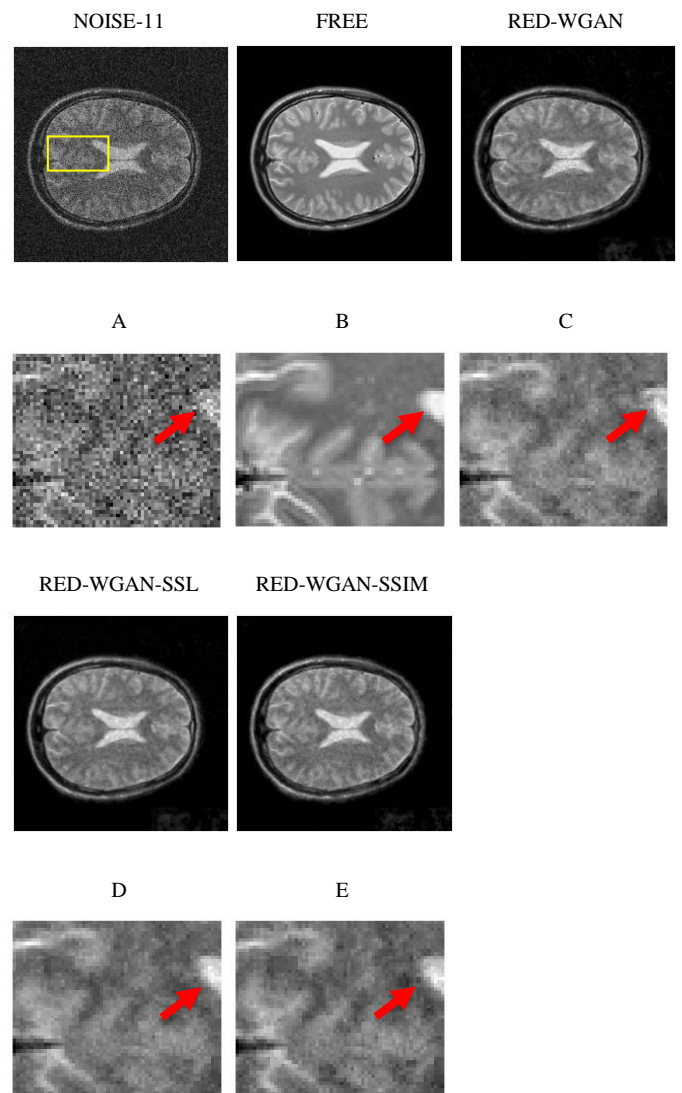


Fig. 2. Denoised PDw example with 11% Rician noise from the testing set at a mini-batch size =110: (A) Noisy image, (B) Ground truth image, (C) RED-WGAN, (D) RED-WGAN-SSL and (E) RED-WGAN-SSIM.

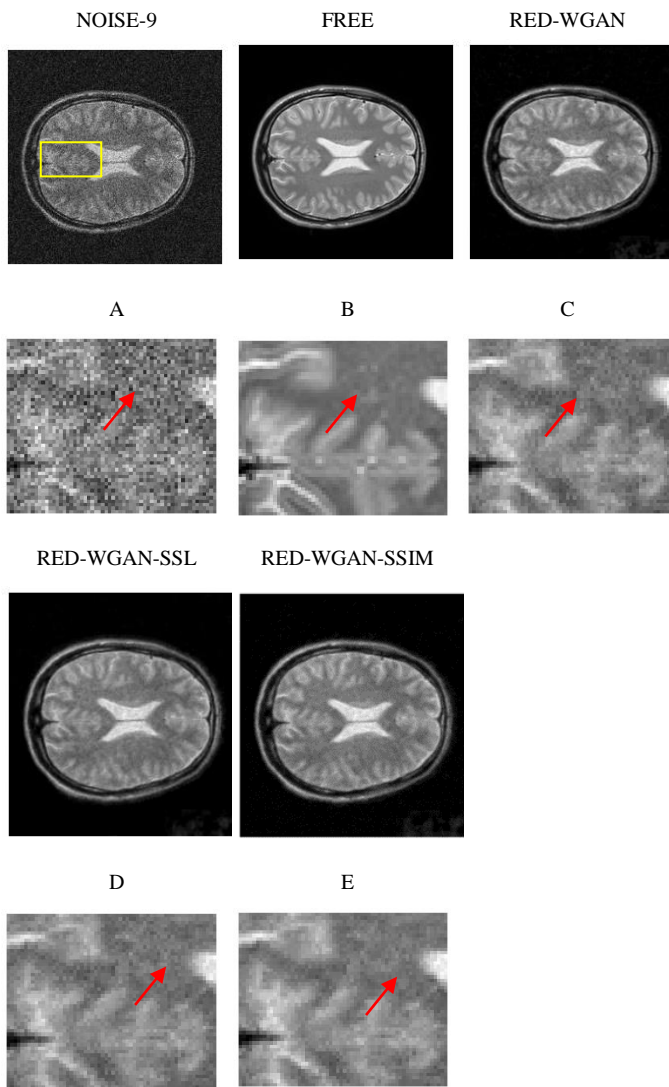


Fig. 3. Denoised PDw example with 9% Rician noise from the testing set at a mini-batch size =110: (A) Noisy image, (B) Ground truth image, (C) RED-WGAN, (D) RED-WGAN-SSL and (E) RED-WGAN-SSIM.

TABLE II. A COMPARISON OF PSNR, SSIM AND RMSE METRICS ON DENOISED PDW EXAMPLE WITH 15% RICIAN NOISE FROM THE TESTING SET AT A MINI-BATCH SIZE =80

		PSNR	41.853721
	Noise -15	SSIM	0.109305
		RMSE	1.135392
	RED-WGAN	PSNR	58.497804
		SSIM	0.755688
		RMSE	0.212351
	RED-WGAN-SSL	PSNR	58.673452
		SSIM	0.781521
		RMSE	0.205079

		PSNR	58.342468
	RED-WGAN-SSIM	SSIM	0.736552
		RMSE	0.212479

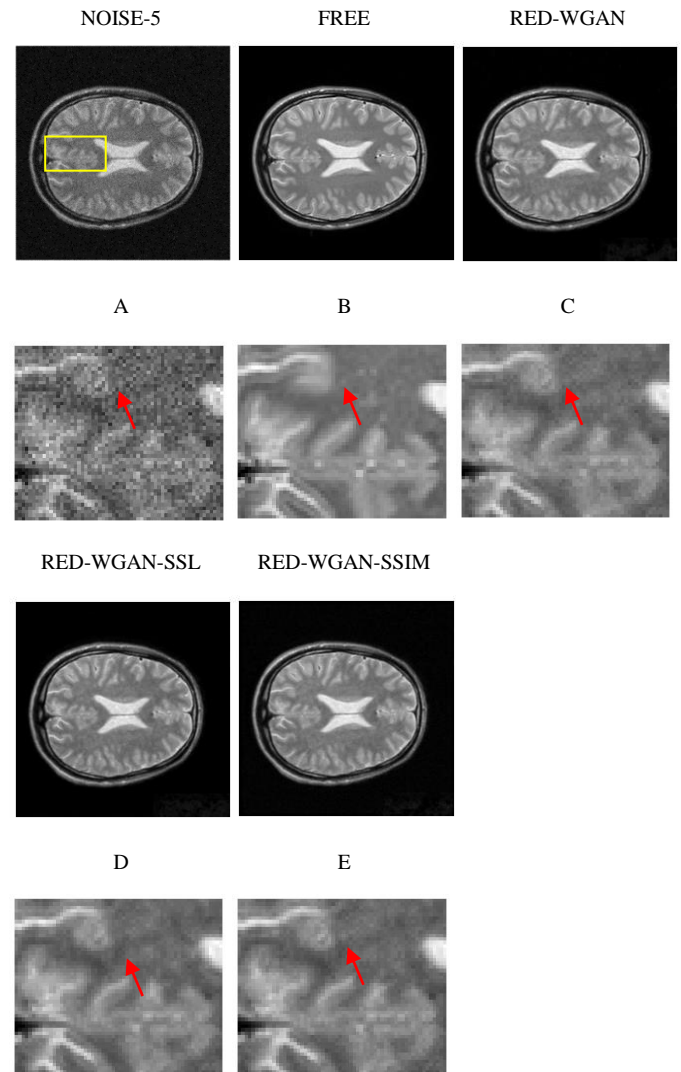


Fig. 4. Denoised PDw example with 5% Rician noise from the testing set at a mini-batch size =110: (A) Noisy image, (B) Ground truth image, (C) RED-WGAN, (D) RED-WGAN-SSL and (E) RED-WGAN-SSIM.

3) Comparison between the results obtained with a mini-batch size = 80 and mini-batch size = 110: Quantitative results of all models at a noise level of 15% with a mini-batch size of 80 are significantly better than those with a mini-batch size of 110 as shown in Table I and Table II. Based on the comparison, we found that the qualitative results with a mini-batch of 80 show that most noise can be effectively removed in most cases, as well as that the structural details are preserved better than the results with a mini-batch of 110 as shown in Fig. 5

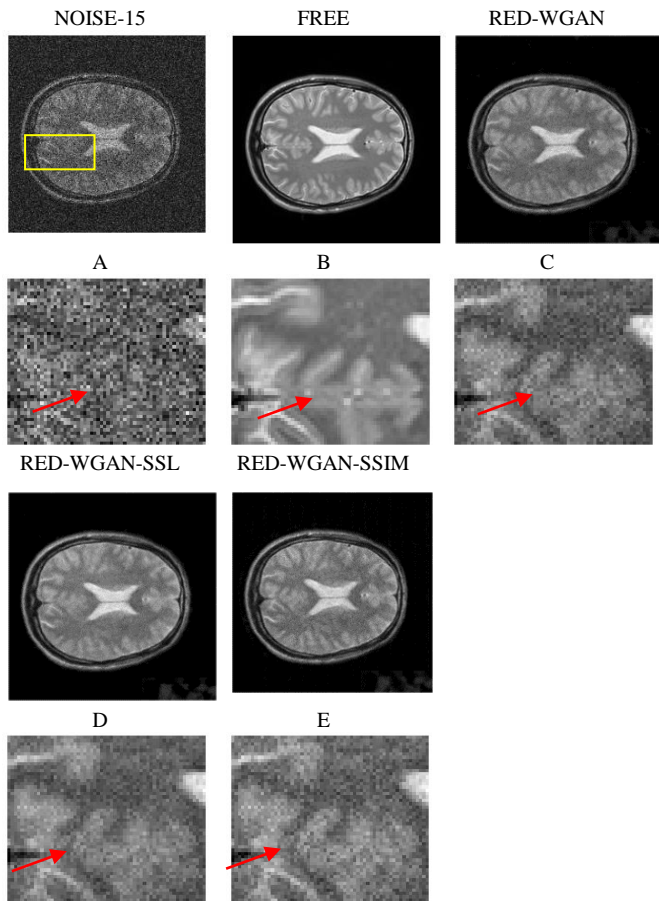


Fig. 5. Denoised PDw example with 15% Rician noise from the testing set at a mini-batch size =80: (A) Noisy image, (B) Ground truth image, (C) RED-WGAN, (D) RED-WGAN-SSL and (E) RED-WGAN-SSIM.

4) *An evaluation of robustness:* For the analysis of the robustness of the RED-WGAN-SSL and RED-WGAN-SSIM

models for various noise levels, RED-WGAN-SSL, RED-WGAN-SSIM and RED-WGAN models were trained with 15% Rician noise, and these three models were then tested with various noise levels which are 5%, 9%, 11%, 15%, and 17% to show how robust they are.

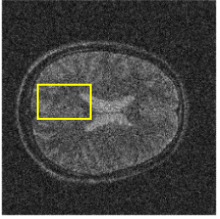
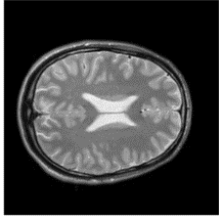
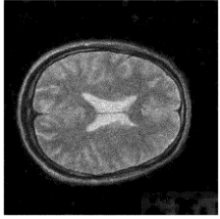

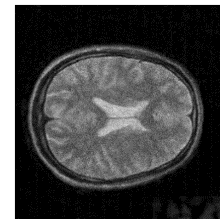
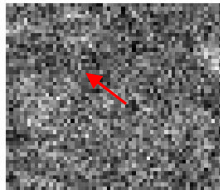
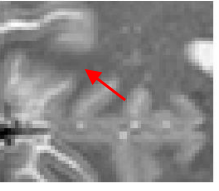
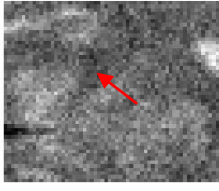
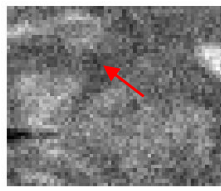
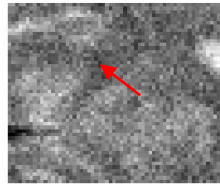
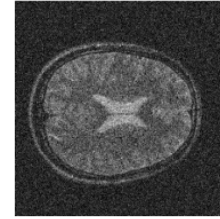
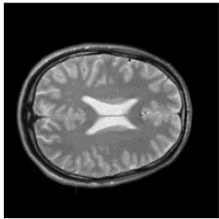
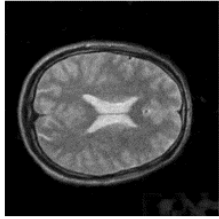
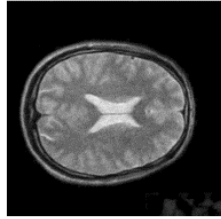
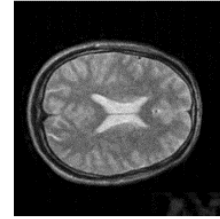
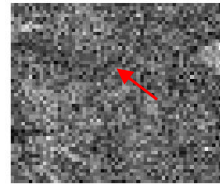
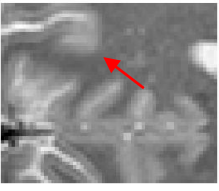
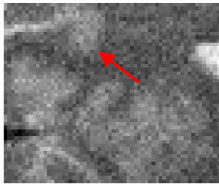
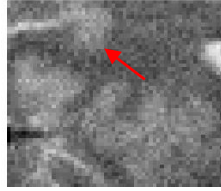
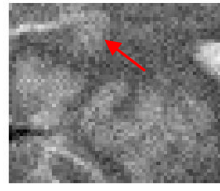
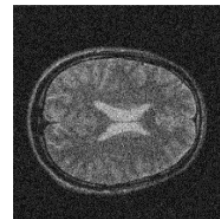
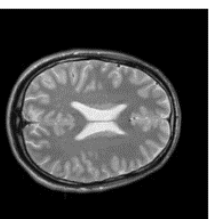
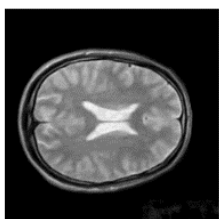
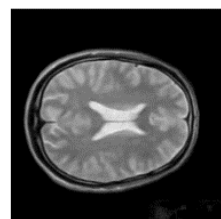
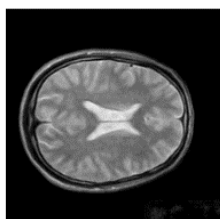
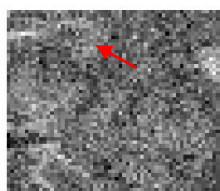

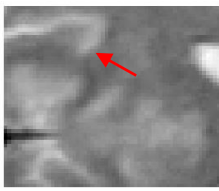
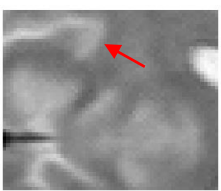
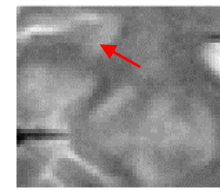
a) *Qualitative Results:* It has been found that the performance of RED-WGAN-SSL is slightly better than that of other models at a higher level than 11% Rician noise. RED-WGAN-SSIM preserved more informative features and provided better visual quality compared to other denoising on the testing set at a lower level than 11% Rician noise, it can reduce the noise and artifacts as indicated by the red arrow in Fig. 6.

The advantage of our proposed models is combining loss functions which are SSL, SSIM, and VGG losses. The VGG loss is used to preserve image style and content after it has been denoised. SSL loss is efficient in extracting structural details and informative features. The SSIM loss generates visually artistic images by using the visible structures in the image. All of these losses help to generate results that can be similar to the original distribution of the data.

b) *Quantitative Results:* A quantitative summary of the results for Fig. 6 is provided in Table III. The RED-WGAN-SSL has better scores when tested at noise levels higher than 11%; it provides good PSNR and SSIM values, which are higher than those of other models as observed in Table III. The RED-WGAN-SSIM has better scores when tested at noise levels less than 11%. As a result, we can take this as evidence that the proposed models are both robust and generalizable. Consequently, we can conclude that the proposed models can denoise MR images with high-quality images and with high structural similarity between the original image and its denoised result.

TABLE III. A COMPARISON OF PSNR, SSIM, AND RMSE MEASURES ON PDW IMAGES WITH DIFFERENT NOISE LEVELS IS SHOWN FROM TOP TO BOTTOM

Noise	RED-WGAN-15	RED-WGAN -SSIM-15	RED-WGAN -SSL-15	Noise
5%	52.446672	53.204429	53.375807	52.93542
	0.306874	0.664977	0.684727	0.665400
	0.374527	0.350747	0.343258	0.361404
9%	46.758666	56.432396	56.666630	56.06631
	0.192677	0.734335	0.745195	0.733152
	0.678883	0.254213	0.247617	0.264126
11%	44.828463	58.067011	58.161415	57.496115
	0.157860	0.763594	0.775136	0.762014
	0.831107	0.214020	0.210872	0.228062
15%	41.853721	58.497804	58.342468	58.673452
	0.109305	0.755688	0.736552	0.781521
	1.135392	0.212351	0.212479	0.205079
17%	40.643732	57.271908	56.900920	57.800380
	0.091793	0.713646	0.644451	0.758627
	1.289173	0.252457	0.257386	0.231039
19%	39.576729	55.302866	54.905874	56.213300
	0.077927	0.626172	0.539672	0.713555
	1.442252	0.315957	0.326715	0.282167

Noise Level	NOISY	FREE	RED-WGAN-15	RED-WGAN-SSL-15	RED-WGAN-SSIM-15
17%					
					
15%					
					
11%					
					

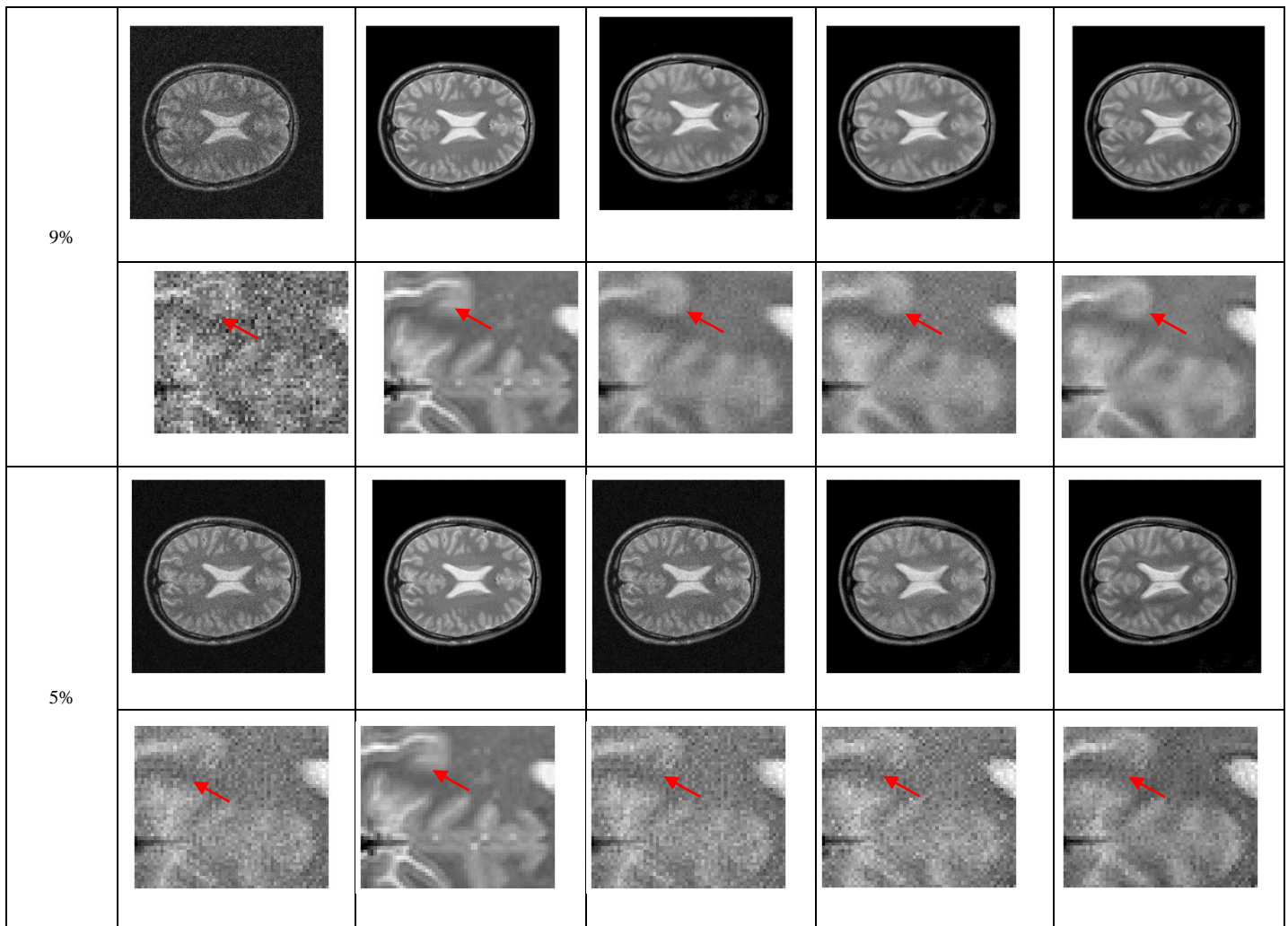


Fig. 6. Denoised PDw example with different levels of Rician noise from the testing set. (A) Noisy image, (B) Ground truth image, (C) RED-WGAN-15, (D) RED-WGAN-SSL-15 and (E) RED-WGAN-SSIM-15.

V. CONCLUSION

The two models RED-WGAN-SSL and RED-WGAN-SSIM models were presented in this paper, which use WGAN to get rip Rician noise from MR images while maintaining structure details. A 3D CNN has been used in these models to process 3D volume data. As well as using the WGAN framework, we introduced an autoencoder generator structure and combined loss functions. We have also improved the performance of our models by adapting the mixture of SSL, SSIM, and VGG loss functions. According to the results of the experiments, the performance of RED-WGAN-SSL and RED-WGAN-SSIM, which are based on the WGAN, SSL, SSIM, and perceptual loss, have been significantly improved both qualitatively and quantitatively. Compared with the RED-WGAN model, they can suppress the noise at the same time as retaining a higher level of detail. A comparison of the results of all models at a noise level of 15% when a mini-batch size = 80 is superior to a mini-batch size = 110. It is interesting to note that the RED-WGAN-SSL scores better metrics on the testing set at noise levels higher than 11%, whereas the RED-WGAN-SSIM scores better metrics on the testing set at noise levels less

than 11%. This leads us to conclude that our proposed models are both robust and generalizable and can therefore be viewed as a strong indication that our work is well worth the effort. A deep learning-based method has a high computational cost. Most of the costs are incurred during the training stage. Although most training is conducted on a GPU, it still takes a long time. In future work, the proposed models will be implemented in T1 and T2 brain image volumes. As well as this, we will apply our denoising methods to a variety of medical images with different types of noise.

REFERENCES

- [1] S. Li, J. Zhou, D. Liang and Q. Liu, "MRI denoising using progressively distribution-based neural network, Magnetic resonance imaging, vol. 71, pp. 55-68, 2020.
- [2] L. Wu, S. Hu and C. Liu, "Denoising of 3D Brain MR Images with Parallel Residual Learning of Convolutional Neural Network Using Global and Local Feature Extraction," Computational Intelligence and Neuroscience, pp. 1-18, 2021.
- [3] M. Ran, J. Hu, Y. Chen, H. Chen, H. Sun, et al., "Denoising of 3D magnetic resonance images using a residual encoder-decoder Wasserstein generative adversarial network. Medical image analysis, vol. 55, pp. 165-180, 2019.

- [4] J. Yang, J. Fan, D. Ai, S. Zhou, S. Tang et al., "Brain MR image denoising for Rician noise using pre-smooth non-local means filter. Biomedical engineering online, vol. 14, no. 1, pp. 1-20, 2015.
- [5] E. Arias Castro and D.L. Donoho "Does median filtering truly preserve edges better than linear filtering?" vol. 37, no. 3, pp. 1172-1206, 2009.
- [6] R.A. Haddad and A.N. Akansu, "A class of fast Gaussian binomial filters for speech and image processing," IEEE Transactions on Signal Processing, vol. 39, no. 3, pp.723-727, 1991.
- [7] J. Chen, J. Benesty, Y. Huang and S. Doclo "New insights into the noise reduction Wiener filter," IEEE Transactions on Audio, Speech and Language Processing, vol. 14, no. 4, pp. 1218-1234, 2006.
- [8] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic Diffusion," IEEE Trans. Pattern Anal. Mach. Intell, vol. 12, no. 7 pp. 629-63, 1990.
- [9] G. Gerig, O. Kubler, R. Kikinis and F.A. Jolesz, "Nonlinear anisotropic filtering of MRI data," IEEE Trans. Med. Imaging, vol. 11, no. 2 pp. 221-232, 1992.
- [10] J. E. Fowler, "The redundant discrete wavelet transform and additive noise," IEEE Signal Processing Letters, vol. 12, no. 9, pp. 629-632, 2005.
- [11] K. Dabov, A. Foi, V. Katkovnik and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Transactions on image processing, vol. 16, no. 8, pp. 2080-2095, 2007.
- [12] X. Zhang, Z. Xu, N. Jia, W. Yang, Q. Feng et al., "Denoising of 3D magnetic resonance images by using higher-order singular value decomposition," Medical image analysis, vol. 19, no. 1, pp.75-86, 2015.
- [13] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P. A. Manzagol, et al., "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," Journal of machine learning research, vol. 11, no. 12, pp. 1-38, 2010.
- [14] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE transactions on image processing, vol. 26, no. 7, pp. 3142-3155, 2017.
- [15] X. Tao, H. Gao, X. Shen, J. Wang and J. Jia, "Scale-recurrent network for deep image deblurring," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8174-8182, 2018.
- [16] C. Dong, C.L. Chen, K. He, X. Tang, "Image super-resolution using deep convolutional networks," IEEE Trans. Pattern Anal. Mach. Intell, vol. 38, no. 2, pp. 295-307, 2015.
- [17] K. G. Lore, A. Akintayo and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," Pattern Recognition, vol. 61, no. 3, pp. 650-662, 2017.
- [18] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," In International conference on machine learning, vol. 37, pp. 448-456, 2015.
- [19] K. Zhang, W. Zuo and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4608-4622, 2018.
- [20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, pp.1-14, 2014, [Online]. Available: <https://arxiv.org/abs/1409.1556>.
- [21] H. Zhao, O. Gallo, I. Frosio and J. Kautz, "Loss functions for image restoration with neural networks," IEEE Transactions on Computational Imaging, vol. 3, no. 1, pp. 47-57, 2016.
- [22] D. Bank, N. Koenigstein and R. Giryes, "Autoencoders," arXiv preprint arXiv:05991, pp. 1-22, 2021.
- [23] D. Jiang, W. Dou, L. Vosters, X. Xu, Y. Sun et al., "Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network," Japanese journal of radiology, vol. 36, pp. 566-574 (2018).
- [24] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778, 2016.
- [25] J.V. Manjón, P. Coupé and A. Buades, "MRI noise estimation and denoising using non-local PCA," Medical image analysis, vol. 22, no. 1, pp. 35-47, 2015.
- [26] P. C. Tripathi and S. Bag, "CNN-DMRI: a convolutional neural network for denoising of magnetic resonance images," Pattern Recognition Letters, vol. 135, no. 1, pp. 57-63, 2020.
- [27] H. Aetesam and S. K. Maji, "Noise-dependent training for deep parallel ensemble denoising in magnetic resonance images," Biomedical Signal Processing and Control, vol. 66, pp. 102405-102415, 2021.
- [28] S. Gregory, H. Cheng, S. Newman and Y. Gan, "HydraNet: a multi-branch convolutional neural network architecture for MRI denoising," In Medical Imaging 2021: Image Processing, vol. 11596, pp. 881-889, 2021.
- [29] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," Magnetic Resonance. Med. vol. 34, no. 6, pp. 910 - 914,1995.
- [30] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125-1134. 2017.
- [31] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, pp. 4681-4690. 2017.
- [32] H. Shi, J. Dong, W. Wang, Y. Qian, and X. Zhang, "SSGAN: Secure Steganography Based on Generative Adversarial Networks," In Pacific Rim Conference on Multimedia, pp. 534-544, 2017.
- [33] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," arXiv preprint arXiv:1511.06434, 2015, [Online]. Available: <https://arxiv.org/abs/1511.06434>.
- [34] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford et al., "Improved techniques for training GANs," Advances in Neural Information Processing Systems, Barcelona, Spain, vol. 29, pp. 1-9, 2016.
- [35] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumouli and A. Courville, "Improved training of Wasserstein GANs," Advances in Neural Information Processing Systems, Long Beach, CA, USA, vol. 30, pp. 5768-5778, 2017.
- [36] C. You, Q. Yang, H. Shan, L. Gjestebj, G. Li et al., "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," IEEE Access, vol. 6, pp. 41839-41855, 2018.
- [37] A. A. Mahmoud, H. A. Sayed and S. S. Mohamed, "Variant wasserstein generative adversarial network applied on low dose ct image denoising," Computers, Materials & Continua, vol. 75, no.2, pp. 4535-4552, 2023.
- [38] <http://brain-development.org/ixi-dataset/>, Accessed 10 March. 2022.
- [39] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi et al., "Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1348-1357, 2018.
- [40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, pp.1-15, 2014, [Online]. Available: <https://arxiv.org/abs/1412.6980>.

A Framework for Patient-Centric Medical Image Management using Blockchain Technology

Abdulaziz Aljaloud

College of Computer Science and Engineering, University of Ha'il, Saudi Arabia

Abstract—In smart systems context, the storage and distribution of health-critical data – medical images, test reports, clinical information etc. that is processed and transmitted via web portal and pervasive devices which requires a secure and efficient management of patients' medical records. The reliance on centralized data centers in the cloud to process, store, and transmit patients' medical records poses some critical challenges including but not limited to operational costs, storage space requirements, and importantly threats and vulnerabilities to the security and privacy of health-critical data. To address these issues, this research proposes a framework and provides a proof-of-the-concept named Patient-Centric Medical Image Management System (PCMIMS). The proposed solution PCMIMS utilizes the Ethereum blockchain and Inter-Planetary File System (IPFS) to enable secure and decentralized storage capabilities that lack in existing solution for patients' medical image management. The PCMIMS design facilitates secure access to Patient-Centric information for health units, patients, medics, and third-party requestors by incorporating the Patient-Centric access control protocol, ensuring privacy and control over medical data. The proposed framework is validated through the deployment of a prototype based on smart contract executed on Ethereum TESTNET blockchain that demonstrates efficiency and feasibility of the solution. Validation results highlight a correlation between (i) number of transactions (i.e., data storage and retrieval), (ii) gas consumption (i.e., energy efficiency), and (iii) data size (volume of Patient-Centric medical images) via repeated trials in Microsoft Windows environment. Validation results also indicate computational efficiency of the solution in terms of processing three most common types of Patient-Centric medical images namely (a) Magnetic resonance imaging (MRI) (b) X-radiation (X-Rays), (c) Computed tomography (CT) scan. This research primarily contributes by designing, implementing, and validating a blockchain based practical solution for efficient and secure management of Patient-Centric medical image management in the context of smart healthcare systems.

Keywords—Smart healthcare; medical imaging; blockchain; ethereum; distributed storage

I. INTRODUCTION

Information and Communication Technologies (ICTs) and infrastructures have brought significant advancements to the healthcare industry – enabling efficient healthcare delivery – to a number of stakeholders such as medical units, patients, and medics [1]. In smart healthcare context, patient information systems have revolutionized the healthcare sector via medical information systems, pervasive technologies, internet of things driven sensors to tackle healthcare challenges effectively and efficiently [2]. Numerous studies have established a strong correlation between storage,

analytics, and transmission of advanced clinical information and communication systems to improve healthcare quality, safety, and patient-centeredness [3]. However, the fragmented nature of health-critical information (i.e., medical images, test reports, clinical information etc.) poses challenges correspond to an effective utilization of personal health information to promote effective and efficient treatment [4]. This fragmentation often results in cumbersome data management and compromises overall patient safety, as health information systems within hospitals, medical clinics, laboratories, and pharmacies remain separate entities that are prone to data security threats or compromised health-critical information [5].

Smart healthcare systems have become increasingly integrated into various aspects of healthcare, including hospital information systems, internet of medical things (IoMTs) and pervasive healthcare technologies [6]. Specifically, as the fundamental unit of smart healthcare, hospital management information systems play a critical role in managing patient data, streamlining workflows, and improving overall healthcare delivery [7]. These systems encompass a range of functionalities, including Electronic Health Records (EHRs), Computerized Physician Order Entry (CPOE), and Picture Archiving and Communication Systems (PACS). PACS has emerged as an essential component of modern radiology departments, enabling the efficient storage, retrieval, and distribution of medical images, including MRI scans [8]. By digitizing and organizing images, PACS eliminates the need for traditional film-based methods and enables rapid access to patient imaging data. It provides radiologists and healthcare professionals with a centralized platform for image interpretation, collaboration, and reporting, leading to improved diagnostic accuracy and patient care outcomes [9].

The integration of ICT in smart healthcare systems, such as hospital information systems and PACS, offers numerous advantages related to cost-efficiency, operational readiness, along with security and privacy of health-critical data [10]. It enables seamless communication and exchange of patient information between various healthcare units (e.g., hospital and labs), medics (e.g., doctors, nurses) and other stakeholders such as patients or healthcare authorities [11]. Such an automated and connected healthcare facilities improve healthcare coordination, faster access to critical information, and enhanced decision-making processes, ultimately benefiting patient outcomes and satisfaction. Despite the outlined benefits as above, the implementation and integration of IT systems, particularly PACS, present various challenges

and such challenges include but are not limited to issues related to interoperability, data security, privacy concerns, and the need for efficient data storage and management solutions [12]. Overcoming these challenges is crucial to ensure the effective utilization of ICT for smart healthcare and maximize its potential for improving patient care [13]. This research aims to address these challenges by proposing a novel approach for enhancing the integration and utilization of ICT systems, specifically focusing on hospital information systems and PACS in the context of MRI imaging. By leveraging existing method and algorithms of blockchain technology, this study aims to optimize data management, improve interoperability, and enhance overall efficiency in healthcare settings.

Context and Challenge(s): Medical images pose a significant challenge in terms of their size and storage requirements, surpassing the capacities typically available on public blockchains [14]. In recent years, blockchain-based solutions have started to gain significant attention with their use-cases and applicability in healthcare data management [15]. In response to the need for decentralized storage, Protocol Labs developed the InterPlanetary File System (IPFS) [16]. IPFS is designed as a distributed web solution that enables the sharing and storage of hypermedia that is managed as peer-to-peer (P2P) file system. It offers an alternative off-chain storage option that can be seamlessly integrated with different blockchain networks, enhancing interoperability [17]. IPFS provides a distributed data access system that offers several advantages, including persistence, improved efficiency, and faster online services. By utilizing content addressing, IPFS ensures that data can be accessed and retrieved from multiple locations, enhancing redundancy and availability [18]. The peer-to-peer nature of IPFS allows for efficient data sharing among participants, reducing the reliance on centralized servers and improving data transfer speeds [8-18]. Moreover, IPFS introduces a more intelligent approach to online services by enabling the creation of decentralized applications (dApps) that leverage its distributed storage capabilities [8-18].

However, despite the advantages offered by distributed storage options, there are notable challenges and barriers when it comes to storing sensitive medical images. One of the primary concerns is ensuring the privacy and security of patient images, while also preventing unauthorized access [21]. The sensitive nature of medical data necessitates robust privacy measures to protect patient confidentiality and complying with necessary regulations including but not limited to Health Insurance Portability and Accountability Act (HIPAA) [22]. Furthermore, a key requirement for a secure data management and its ability to handle large volumes of data across various stakeholders, including medics, health units, patients, and institutions [23].

Solution Overview: To address these challenges, we design and implement a proof-of-the-concept for a distributed framework called a Patient-Centric image management (PCMIMS). PCMIMS is a blockchain-based solution that is architected to enable secure and private management of Patient-Centric medical images within an open distributed network. We elaborate on a step-wise and incremental design,

implementation and validation of the PCMIMS solution in dedicated sections and outline the key contributions and salient features of the proposed solution below. The implications of this research and proposed solution aims to complement the research and development on exploiting blockchain technologies synergized with the IPFS to enable secure management of health-critical data. The proposed solution PCMIMS in terms of a framework, its implemented prototype, and validations could help researchers and developers to incrementally design and develop a blockchain-based Patient-Centric medical image management in a secure and efficient way. Specifically, the solution aims to improve state-of-the-art by offering:

- Blockchain-based framework (i.e., PCMIMS) that provides a structural representation of the overall solution that sketches a blue-print and guides software designers and developers about how the system operates and facilitates effective healthcare management.
- Smart contract-based implementation of Patient-Centric access to PCMIMS that ensures controlled access to medical image data for stakeholders, i.e., medics and patients. The implemented prototype employs specific functions to enable data transfer via Ethereum blockchain and establishes access privileges between relevant parties. This enhances privacy and security in the management of medical data.
- Experimental validation of the functionality of the proposed solution PCMIMS through test cases, focused on key performance metrics. These metrics include (i) number of transactions (i.e., data storage and retrieval), (ii) gas consumption (i.e., energy efficiency), and (iii) data size (volume of Patient-Centric medical images) via repeated trials in Microsoft Windows environment. This evaluation ensures the effectiveness and efficiency of the proposed framework in handling medical data and provides insights for future enhancements.

Structure of the paper: Section II provides context and background of the proposed research. Section III discusses the most relevant existing research. Section IV presents research method and motivating scenario. Section V discusses prototype-based implementation of the proposed solution. Section VI presents results of the solution validation. Section VII concludes the paper and summarizes the key findings and contributions of this research.

II. RESEARCH CONTEXT AND BACKGROUND

In this section, we provide a broader perspective and overall context of the proposed research. The context provides necessary background information in terms of some core concepts of healthcare management, distributed system, and healthcare data management. We have introduced the concepts and terminologies introduced in this section that are used throughout the paper to explain the technical concepts. Fig. 1 illustrates the core concepts and complements the theoretical description presented in this section.

A. Digital Healthcare Systems

Healthcare systems heavily rely on medical image systems for diagnosis, treatment, and monitoring of various medical conditions. These systems encompass the storage, retrieval, and analysis of medical images, specifically addressing X-rays, CT scans, MRI scans, and ultrasound images. The effective utilization of medical image systems plays a vital role in improving patient outcomes and optimizing healthcare delivery. However, the widespread adoption and integration of medical image systems in healthcare settings pose numerous challenges. One significant challenge is the sheer volume and complexity of medical images generated daily, leading to storage and management issues [12]. The exponential growth in medical image data requires robust infrastructure and storage solutions to ensure efficient access, retrieval, and secure archiving.

B. Medical Image Data

Another critical challenge lies in the interoperability and integration of medical image systems within the broader healthcare ecosystem [13]. Healthcare providers, hospitals, clinics, and other stakeholders often employ different systems and technologies, resulting in fragmented data silos. This fragmentation hinders seamless data exchange, collaboration, and comprehensive patient care. Moreover, privacy and security concerns are paramount in healthcare systems, especially when it comes to sensitive patient information contained within medical images [14]. The protection of patient privacy, compliance with regulatory requirements (e.g., HIPAA), and safeguarding against unauthorized access are essential considerations in the design, development, and operational context of medical image systems. Addressing these challenges requires innovative approaches and technologies to enhance the efficiency, interoperability, and security of medical image systems. Solutions such as distributed storage, blockchain technology, and advanced data analytics have shown promise in overcoming these hurdles [15, 16, 17]. By leveraging these advancements, healthcare organizations can optimize the management and utilization of medical images, leading to improved patient care, streamlined workflows, and enhanced decision-making processes.

Fig. 1 provides a comprehensive illustration of medical image-based data management in a distributed environment. The system ensures that user authorization is a prerequisite for accessing its functionalities. The user base comprises medics, patients, and individuals undergoing lab tests. In the patient registration process, users are required to request permission to utilize the system. Once the authorization is granted, users who have undergone lab tests can proceed to upload their test results. The MRI and radiologist specific data can be accessed via a designated web portal. It is important to note that access to the report's data is limited only to patients and doctors who have been specifically granted permission. A rigorous access control mechanism and implementation ensures that health-critical data remains private and secure and is only accessible to authorized individuals involved in patient care and treatment. The system's design and implementation prioritize data privacy and confidentiality while enabling efficient and

secure sharing of medical information within a trusted and authorized user community.

III. RELATED WORK

This section reviews the most relevant existing research in terms of discussing state-of-the-art on (i) medical image data management in centralized system in Section III A and medical image data in distributed and peer-to-peer systems in Section III B. A critical review of the most relevant existing work helps us to comparatively analyze the scope of proposed solution and justify its contributions. Table I provides the comparative analysis and act as a catalogue to distinguish between most relevant existing works and proposed solution PCMIMS.

A. Medical Imaging Data in Centralised Systems

The management of medical health records is a critical aspect of healthcare systems. Traditional paper-based records have been gradually replaced by electronic health records (EHRs), which offer numerous advantages in terms of accessibility, interoperability, and data sharing [18]. EHR systems have transformed the way patient information is stored and accessed, enabling healthcare providers to have comprehensive and real-time access to patient data. Advanced medical imaging techniques, such as CT scans, MRI scans, and X-rays, play a pivotal role in diagnosing and monitoring various medical conditions. These imaging modalities generate large volumes of data, and efficient storage and retrieval systems are essential for their effective utilization in healthcare [19]. Imaging systems need to handle the complexities of different image formats, metadata, and associated clinical information. The management of medical images involves storing, organizing, and accessing various types of images, including CT scans, MRI scans, and X-rays. With the increasing volume and complexity of medical images, there is a growing need for efficient storage, retrieval, and analysis solutions [20]. Centralized and decentralized approaches have been explored to address the challenges of managing and sharing medical images.

Centralized systems, such as cloud-based storage and data centers, have been widely used in healthcare for storing and managing medical images. These systems offer centralized control, efficient storage, and accessibility but raise concerns regarding data security, privacy, and single points of failure [21].

B. Medical Imaging in Peer-to-Peer Distributed Systems

In contrast to the centralized management of medical images, decentralized approaches, such as distributed file systems and blockchain technology, have gained attention as potential solutions for medical image management [22].

Distributed file systems, like the InterPlanetary File System (IPFS), offer a decentralized architecture for efficient and secure storage and sharing of medical images [23]. Blockchain technology provides a distributed and immutable ledger that ensures data integrity, transparency, and privacy in medical image systems [24].

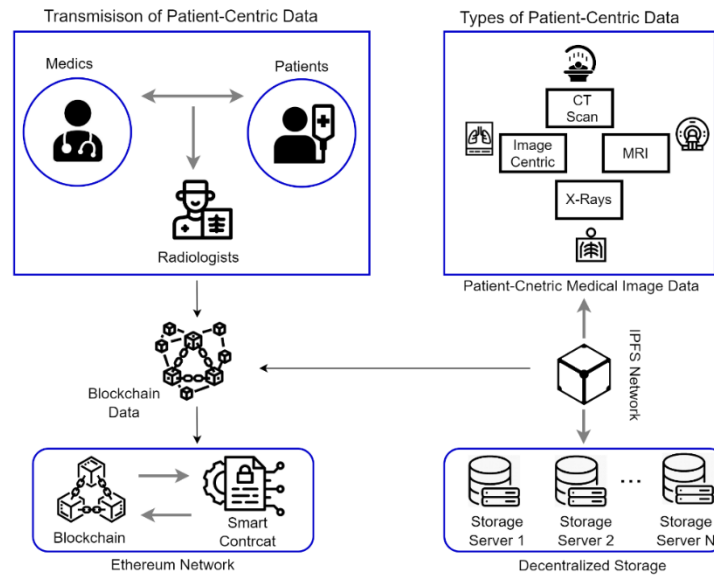


Fig. 1. Context: decentralized management of medical image data.

Researchers have explored the integration of centralized and decentralized approaches to strike a balance between efficiency, scalability, security, and privacy in medical image systems [25]. These studies aim to leverage the benefits of decentralized systems while addressing the challenges associated with data management, privacy, and interoperability.

Fig. 1 provides a comprehensive illustration of the entire system, depicting each step and process involved. The system ensures that user authorization is a prerequisite for accessing its functionalities. The user base comprises the medics (doctors, nurses, etc.), patients, and individuals undergoing lab tests. During the patient registration process, the users are required to request permission to utilize the system. Once the authorization is granted, users who have undergone lab tests can proceed to upload their test results and MRI radiologist image data through the designated website. It is vital to mention that access to the report's data is limited only to patients and doctors who have been specifically granted permission. This strict access control mechanism ensures that sensitive medical information remains secure and is only accessible to authorized individuals involved in patient care and treatment. The system's design and implementation prioritize data privacy and confidentiality while enabling efficient and secure sharing of medical information within a trusted and authorized user community.

Comparative Analysis and Summary: Table I provides a concise comparison of the proposed framework with other existing blockchain-based medical imaging and health data management systems. The table highlights the advantages offered by the PCMIM system over its competitors. It is evident that the PCMIM system outperforms the alternatives in terms of various benefits and features.

The table serves as a valuable reference, showcasing the superiority of the PCMIM system in comparison to the

existing alternatives, thereby reinforcing its potential as an innovative and promising solution for effective medical health record management.

TABLE I. COMPARISON BETWEEN THE EXISTING AND PROPOSED PCMIMS SYSTEM

Scheme	[3]	[12]	PCMIMS System
Source Data Storage	Dedicated	Server	Immutable IPFS Storage
Encryption	Symmetric Encryption	✓	Asymmetric Encryption
Server Attack Resistance	×	×	✓
Database Management	Centralized	Centralized	Decentralized
Smart-Contract	×	✓	✓
Data-Access	×	×	✓

IV. RESEARCH CONTEXT AND BACKGROUND

This section presents the research method that is employed to conduct this research study, outlining the design specifics of the proposed solution. Fig. 2 provides an overview of the research methodology, which comprises four distinct stages. These stages follow an incremental approach, allowing for the assessment, development, and validation of the solution.

A. Steps of Research Method

The research methodology followed a structured process consisting of four distinct steps.

- Phase I: Literature Review involved a thorough examination of a diverse range of literature sources, including peer-reviewed research articles, technological road maps, and technical reports, to identify existing solutions and their limitations.

```
01: function AddImageCentric(string memory patientUserId, uint appointment_id,  
02: string memory description, string memory filehash) public{  
03: imgCount ++;  
04: GetImageCentric_Id [_appointment_id] = ImageCentric(imgCount, appointment_id, _description, _filehash, now);  
05: PatientRecordAccess [_patientUserId][_appointment_id] = ImageCentric (imgCount, appointment_id, _description, filehash, now);  
06: emit ImageCentricCreated(imgCount, appointment_id, _description, filehash, now);  
07: }
```

Listing 1. Code snippet for adding image to the chain.

To gather the most relevant papers in the field, a systematic literature review was conducted. This comprehensive review of existing research and development solutions helped establish the research scope and streamline the necessary solutions.

- Phase II: System Design focuses on architectural design of the proposed solution. Here, the proposed solution was meticulously modeled and simulated in accordance with established standards, such as the ISO/IEC/IEEE 42010:2011 standard [26].
- Phase III Algorithmic Implementation encompassed the implementation of algorithms, which involved executing the solution through computational and storage-intensive phases. The solution was broken down into modular algorithms that could be customized by users with specific inputs, adhering to executable standards and underlying source code.
- Phase IV Algorithmic Evaluation steps revolve around validating the solution by assessing its functionality and quality using the ISO/IEC-9126 model [26]. Well-established assessment metrics were employed to evaluate the system's usability and efficiency. While the first two steps required manual effort and human decision-making, the latter two phases involved human interaction and tool support. The validation phase provided insights for potential algorithm enhancements to improve efficiency or modify functionality based on the evaluation results. Overall, this methodology, as depicted in Fig. 2, ensured a comprehensive and systematic approach to solution development and evaluation.

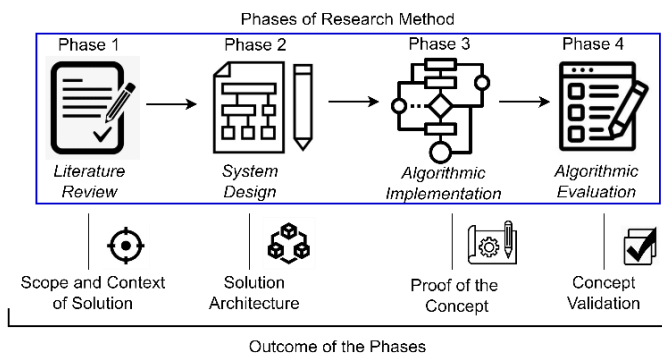


Fig. 2. An overview of four phases of research method.

B. Motivating Scenario

Fig. 3 illustrates a comprehensive process involved in storing and managing medical imaging and test reports within the proposed framework. The medical images are managed and stored via IPFS by the radiologist, who is responsible for the image acquisition and interpretation within the lab. The radiologist uploads the medical image data to IPFS, and in return, receives a unique hash key associated with the stored image. This hash key, along with other necessary information, is then linked and stored within the blockchain. It is important to note that the processes for medical image uploading are distinct from each other. The file hash of the medical image is then linked to the relevant details and stored within the blockchain. This systematic approach ensures the secure and efficient storage of medical images while maintaining the integrity and traceability of the data within the proposed framework.

The process of medical image data sharing is triggered via the creation of metadata for the original data file, which includes essential information such as the file's name, type, description, and size (Line 1 – Line 2) in Listing 1. Once the metadata is complete, it is uploaded to the IPFS along with the corresponding data file. To achieve this, a smart contract is created for the storage of the data in the blockchain by means of a function named 'AddImageCentric' (Line 4) with required parameters (Line 1). In order to streamline the data mapping process, two separate mappings are employed. The first mapping enables retrieval of a comprehensive list of all blood tests that are linked to a patient's appointment ID and the second mapping facilitates access to data specific to individual patients based on two parameters namely patient ID and appointment ID. To provide a seamless user experience and enhance data traceability, the 'ImageCentricCreated' event is triggered, signaling the successful creation of the image-centric data entry. This robust and organized approach ensures efficient data sharing and retrieval within the proposed framework while leveraging the advantages of blockchain technology and IPFS.

Blockchain-based management and IPFS based transmission of data ensures secure and transparent storage and transfer of medical image-centric data. Fig. 4 depicts two distinct categories of medical data uploading within the Decentralized Application (DApp) framework.

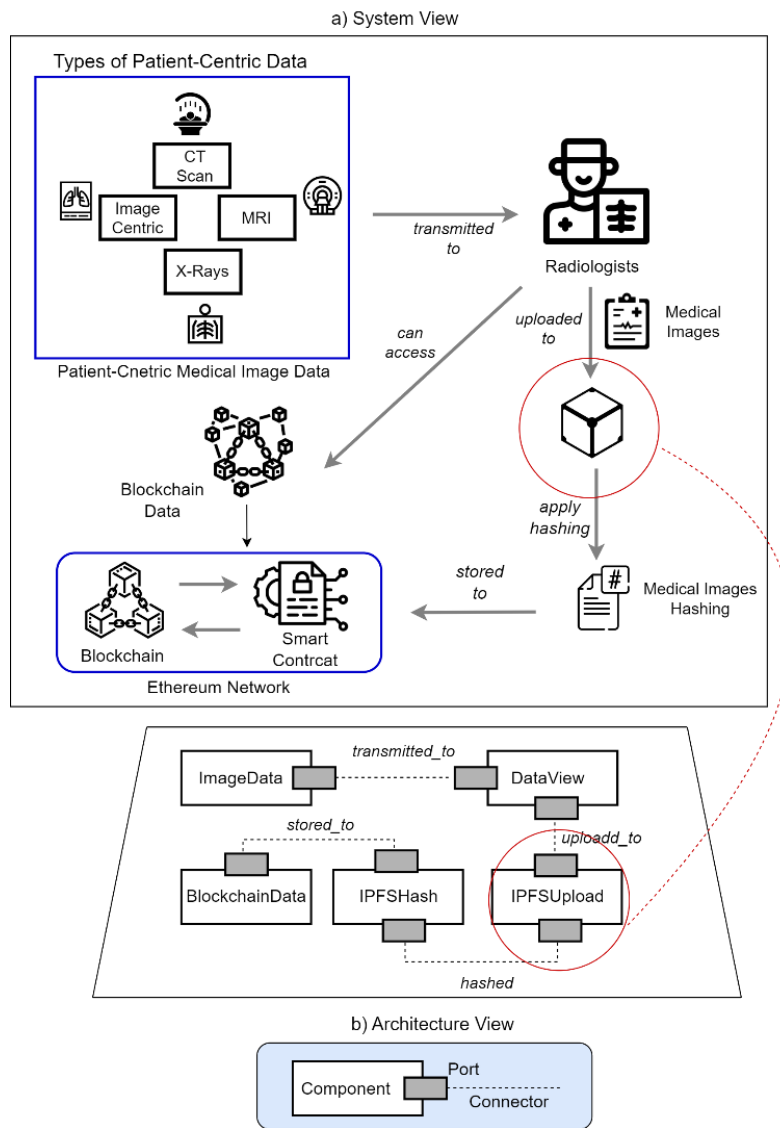


Fig. 3. Patient-Centric data storage process.

(a) Overall System View, b) Architecture View

It pertains to medical images, which are connected to other examinations and tests by a radiologist. The radiologist utilizes the DApp system to upload the medical images to IPFS. Alongside the image upload, the radiologist provides the remaining essential details that include patient ID and appointment ID that are stored and retrieved via blockchain. This integrated approach ensures that medical images are seamlessly linked to the relevant patient and appointment data, enabling efficient retrieval and analysis.

The service cycle depicted in Fig. 4 continues to repeat itself at fixed intervals or auto triggered via receipt of medical image data, provided by the server. The dashboard serves as a platform that offers access to both doctors and patients acting as a web-accessible portal, enabling PCMIMS users to access, view, customize and download the medical images over the network. The web portal provides a convenient means of retrieving the health critical data (medical imaging, test

reports etc.), with data retrieval based on customized parameters such as patient ID and appointment ID mappings [27] [28].

V. IMPLEMENTATION OF THE ALGORITHMS AS SOLUTION PROTOTYPE

This section discusses the implementation details in terms of the developed algorithms that provide a proof-of-the-concept prototype in Section VA. The discussion on algorithms is complemented with the details of tools and technologies for implementing the algorithms in Section V B. A precise discussion of the tools and technologies used for the implementation can help designers and developers to gain insight into the software tools that can be used to implement the solution [29].

A. Algorithmic View of the Developed Solution

The algorithms provide a modular implementation of the designed solution as in Fig. 3. Each of the two algorithms presented below are detailed in terms of their *inputs*, *processing*, and *outputs*, each elaborated below.

Algorithm 1 shows a step-wise and incremental this section demonstrates and describes the uploading feature for medical data. The technique specifically focuses on uploading medical image data to the IPFS and also applies hashing to the medical image data file for an extra security feature before its representation as a smart contract and storage in the blockchain. The uploaded data is associated with a number of parameters that include but are not limited to user and appointment identities (UserID, Appointment ID) along with description (Description), and data of the appointment (Date).

Algorithm 1 Decentralizing Medical Image

Parameters:

U_ID: User ID
A_ID: Appointment ID
Des: Description
Img = Image

```
1: Income: U_ID, A_ID, Des, Img
2: Outcome: O //Returning Result
3: procedure DATACENTRICMODULE // Function
4: if U_ID is Valid then //Decentralizing Image
5: File_Stream ← File(Img) //Get File stream FS
6: File_Buffer ← Buffer.form (File_Stream) //Convert
7: File_Hash ← IPFS.Add (File_Buffer)
8: O ← SBC(U_ID, A_ID, Description, File_Hash) // Store
Image hash and Data to Blockchain
9: end if
10: end procedure
```

- **Income:** The input to the algorithm involves mapping the parameters with a unique hash key of the medical image file. This mapping process occurs as part of the input. The input parameters include User ID, Appointment ID, and Description of the appointment and Medical Image. Combinedly, these parameters provide information about an individual user and their medical related information. The parameters provide customization and parametrized input to the solution.
- **Processing:** During the processing stage, the medical image data is being retrieved and processed for its conversion into a buffer package. The buffered data, comprising of medical image data is stored as IPFS as a medical data file that in turn hashes the data and generates a hashed key for data access. The additional parameters that include user and appointment identities (UserID, Appointment ID) along with description (Description), and data of the appointment (Date) are associated to the hash key. These parameters are subsequently are part of the smart contract that are privately stored on the blockchain.

- **Outcome:** Finally, the output stage involves managing and storing the processed medical image data in the blockchain. The mapped date, along with the associated parameters, is stored as the output of the algorithm, ensuring the secure and traceable storage of the uploaded medical data.

Algorithm 2 focuses on validating data accessing capabilities. The algorithm's primary purpose is to access the data in a secure way and make it accessible to authorized entities. By utilizing this algorithm, users can access specific data stored on the blockchain, driven by customized parameters. Notably, there are multiple types of data access available within the algorithm's processing stage. As an example, a user can access their data via user and appointment identities (UserID, Appointment ID) along with description (Description), and data of the appointment (Date).

Algorithm 2 Presentation Layer

Parameters:

U_ID: User ID
A_ID: Appointment ID
U_Type: User Type
T_Type= Test Type

```
1: Income: U_ID, A_ID, U_Type, T_Type
2: Outcome: O
3: procedure Interface_Module
4: if U_Type == D then //Image Access
5:  $\epsilon$  ← Get_Report(T_Type) // Test Type
6: end if
7: R ← Update_Dashboard( $\epsilon$ ) //Data Dashboard
8: end procedure
```

- **Income:** During the input stage of the algorithm, the settings for data access are mapped, enabling the algorithm to determine the appropriate parameters for retrieval. The input parameters include User ID, Appointment ID, User Types and the medical Test Type. Combinedly, these parameters provide information about an individual user and their medical test being performed.
- **Processing:** The processing stage of the algorithm revolves around granting various forms of data access. This includes allowing users to access their data based on the mapping between their user and appointment identities (UserID, Appointment ID). Moreover, the medics such as the doctors or nurses can readily access the medical image data and reports by utilizing the appointment ID associated with the user (UserID).
- **Outcome:** As a result of the algorithm's execution, the output consists of publicly accessible data that has been mapped according to the defined parameters. This ensures that authorized users, such as patients and doctors, can retrieve the relevant data from the blockchain.

B. Tools and Technologies for Algorithmic Implementation

In this section, we present a concise overview of the complementary tools and technologies employed in the proposed solution, aiming to enhance the reader's comprehension of the underlying technology.

The combination of stacked tools and technologies, as depicted in Fig. 4, offers a comprehensive solution. For instance, if a radiologist accesses the portal, they can securely upload encrypted medical image files to the IPFS platform, receiving a hash key for subsequent validation of the proposed solution [30] [31].

VI. VALIDATION AND RESULTS

After presenting the implementation of the solution, we now discuss details on validation of the solution. To discuss the validation results objectively, we present the evaluation environment in Section VI A. We then elaborate on validation of the solution based on correlation between the number of transactions, gas consumption, and block data size. Finally, we also present computational efficiency of the solution in terms of processing three most common types of Patient-Centric medical images namely (a) Magnetic resonance imaging (MRI) (b) X-radiation (X-Rays), (c) Computed tomography (CT) scan.

A. Setting up the Evaluation Environment

The evaluation environment was carefully designed to include a combination of hardware and software resources that were essential for effectively executing and monitoring the different phases and outcomes of the solution. For the hardware-based evaluation experiments, a Windows platform was utilized, featuring a powerful core i7 processor and 16 GB of RAM. This robust hardware configuration provided the necessary computing power and memory capacity to handle the demanding tasks involved in the evaluation process. On the software front, a set of evaluation scripts was developed using NodeJS and ReactJS, two widely used technologies in the software industry. These scripts were executed within the popular integrated development environment, Visual Studio Code, allowing for efficient code development and testing. The scripts were specifically designed to automate system testing and incorporated various libraries, such as React, Web3, and IPFS.HTTP. These libraries provided essential functionalities and streamlined the evaluation process.

To thoroughly analyze the CPU consumption during the evaluation, a specialized JavaScript performance library script was employed. This script enabled detailed assessment of CPU usage during critical tasks, including the uploading of medical image data to IPFS, storing it securely on the blockchain, and retrieving data from the blockchain. This analysis provided valuable insights into the system's performance and efficiency.

Furthermore, the evaluation environment included the utilization of the Ganache suite, a comprehensive toolset for establishing a local Ethereum blockchain environment. The Ganache suite allowed for the creation of a personalized blockchain network, facilitating testing, command execution,

and observation of the blockchain's state. To enable connectivity with the distributed web and interact with the Ethereum network, the Metamask extension was employed within the browser. This extension seamlessly connected local Ethereum accounts to the Ganache suite, enabling smooth and reliable system functions while considering gas transaction costs.

B. Discussion I - Data Transmission and Gas Consumption

In our suggested solution, we specified the cost of executing the contract migration, which was presented in Table II. This cost was denominated in Ether, and the amount of gas spent during the execution was recorded. To determine the value of Ether, the amount of gas utilized was multiplied by the current gas price. It is important to highlight that the gas price can fluctuate in response to network dynamics and changes in the value of Ether. This dynamic adjustment ensures that the gas cost accurately reflects the ongoing computational expenses within the system, providing a fair and efficient pricing mechanism. During the evaluation process, one of the crucial parameters tested was the time required for users to upload and store data to IPFS and the blockchain ledger. This test parameter encompassed the overall time spent on data. The findings of a series of experiments conducted on an average data size are depicted in Fig. 4. This indicates that the suggested methodology is capable of handling larger data sizes without incurring a notable impact on fuel consumption.

TABLE II. SUMMARY OF THE EXECUTION COST ANALYSIS

Parameter Serial	Computation Efficiency (Execution Time)	Energy Efficiency (Gas Utilisation)	Ether Cost (Overhead)
1	Contract Initialisation	225 (in thousands)	0.058
2	Contract Instantiation	286	0.053
3	Contract Call Initiation	42	0.045
4	Contract Migration	27	0.084
Average		145	0.061

These results validate the effectiveness of the proposed solution in efficiently managing the uploading and storing of medical data to IPFS and the blockchain ledger, even as the data size increases. However, it is important to note that despite the increase in data size, the suggested methodology for uploading medical data to IPFS showcased consistent and efficient fuel usage. The difference in fuel consumption between uploading 450-byte data and 1000-byte data was not found to be significantly significant. This indicates that the suggested methodology is capable of handling larger data sizes without incurring a notable impact on fuel consumption. These results validate the effectiveness of the proposed solution in efficiently managing the uploading and storing of medical data to IPFS and the blockchain ledger, even as the data size increases. The suggested methodology ensures a reliable and streamlined process, enabling users to securely upload and access their medical data with minimal impact on fuel consumption.

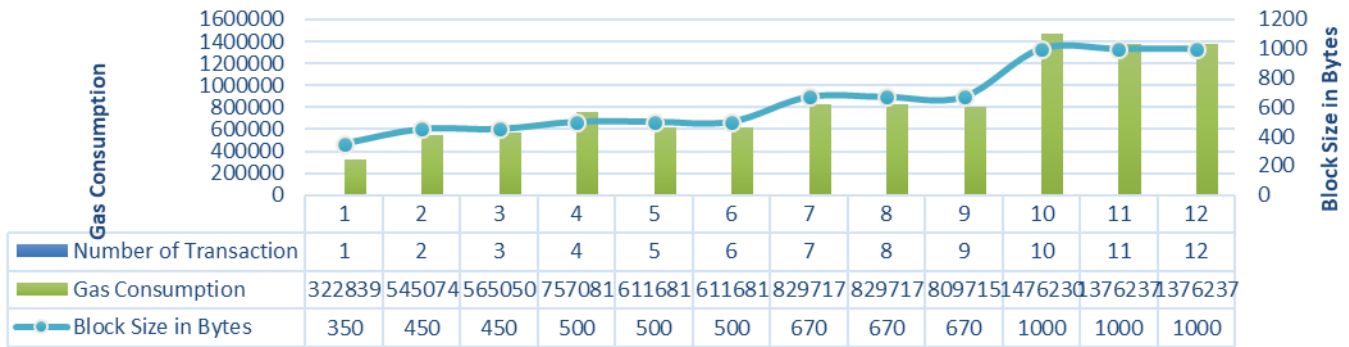


Fig. 4. Gas consumption calculated based on the size of block and number of transactions.

C. Discussion II – Query Response Time

When it comes to storing medical images in IPFS and recording relevant information in the blockchain, efficient data querying becomes a crucial aspect. The query response time plays a vital role in assessing the solution's effectiveness in storing and retrieving data from the blockchain. Fig. 5 provides a visual representation of the execution time involved in accessing the data within this context. The data can be classified into two distinct types. The first type pertains to medical images that are fetched from IPFS using their respective file hashes. This approach ensures a decentralized and distributed storage mechanism for medical images. However, in the case of example number six, it is observed that the execution time is and review, collectively referred to as data uploading and accessing time. The findings of a series of experiments conducted on an average data size are depicted in Fig. 5.

However, it is important to note that despite the increase in data size, the suggested methodology for uploading medical data to IPFS showcased consistent and efficient fuel usage.

The difference in fuel consumption between uploading 450-byte data and 1000-byte data was not found to be significantly significant. However, in the case of example number six, it is observed that the execution time is relatively high compared to other cases. This can be primarily attributed to the larger file size of the medical image associated with this particular scenario. As the file size of the medical image increases, it naturally requires more time for display or download from IPFS. The larger data volume associated with the larger file size contributes to an extended retrieval time. This phenomenon aligns with the intuitive understanding that larger files necessitate more data to be transferred, resulting in a slightly longer execution time. This observation highlights the impact of file size on the retrieval process and emphasizes the need to consider the trade-off between file size and retrieval time when dealing with medical images stored in IPFS. Finding the right balance between image resolution and file size can play a significant role in optimizing the execution time for accessing and displaying medical images in the given solution.

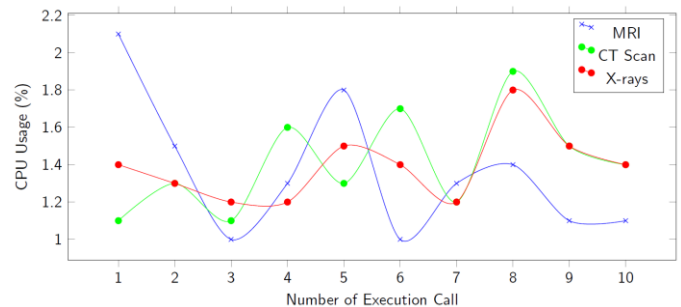


Fig. 5. Computation efficiency and function execution to store the data in IPFS and blockchain.

VII. CONCLUSION AND FUTURE WORK

Decision support and intelligence in healthcare care systems can greatly enhance the cost and efficiency of the healthcare systems. More specifically, the intelligence embedded in electronic healthcare systems relies on efficient processing of medical health data. However, the dispersion of these images across multiple systems poses a challenge for effective and integrated healthcare. Additionally, centralized hosting solutions for picture data, such as cloud-based platforms, can be vulnerable to security breaches. Recognizing the distributed nature of healthcare, there has been increasing attention towards decentralized designs and system interoperability. This research design and implements a proof-of-the-concept for a decentralized Patient-Centric Medical Image Management System (PCMIMS). Built upon the Ethereum blockchain and IPFS, this framework aims to address the storage and distribution challenges associated with medical images. PCMIMS empowers medical images users with secure and private control over their medical images and test reports (such as blood reports, lipid reports) through a DApp web application portal.

1) *Evaluation of the proposed solution:* To evaluate the efficiency, feasibility, and practicality of the proposed scheme, we conducted an experimental implementation. The suggested system not only facilitates the sharing of medical images but also provides patients with access to an immutable medical

database, thereby enhancing efficiency, data provenance, and effective auditing. By employing a decentralized data storage and exchange mechanism, the need for intermediaries or administrative entities is eliminated. Overall, the PCMIMS system presents a unique solution that prioritizes patient rights and control over their medical data. Through decentralization and the utilization of blockchain and IPFS technologies, this system offers enhanced security, privacy, and accessibility in the management of medical images and test reports.

2) *Vision for future research*: In the current scope of the proposed solution, there is a need for future research to increase the pervasiveness of the proposed solution and rigor of its validation. More specifically, the future work is focused on:

- Developing, deploying, and validating the proposed solution in mobile health (mHealth) context. Mobile devices can offer pervasive healthcare but pose significant challenges related to resource poverty of mobile devices along with security and privacy of health critical data.
- Evaluating the proposed solution via survey or trials with the domain experts, i.e., medics and healthcare professionals. An incorporation of the domain experts' feedback can help us to analyse and improve the solution based on practitioner' perspective.

3) *Implications for research and practices*: The implications can be attributed to the relevance and benefits of the proposed solution in an academic and industrial context. Blockchain-based framework (i.e., PCMIMS) that provides a structural representation of the overall solution that sketches a blue-print and guides software designers and developers about how the system operates and facilitates effective healthcare management. Academic researchers can exploit architectural design to design and develop emerging solution for blockchain systems in healthcare context. Practitioners can rely on the algorithms and relevant tools to develop solutions that rely on blockchain solutions for managing electronic healthcare systems.

REFERENCES

- [1] Aljedaani, Bakheet, Aakash Ahmad, Mansoorah Zahedi, and M. Ali Babar. "End-users' knowledge and perception about security of clinical mobile health apps: A case study with two Saudi Arabian mHealth providers." *Journal of Systems and Software* 195 (2023): 111519.
- [2] Aljedaani, Bakheet, Aakash Ahmad, Mansoorah Zahedi, and Muhammad Ali Babar. "An empirical study on secure usage of mobile health apps: The attack simulation approach." *Information and Software Technology* 163 (2023): 107285.
- [3] Casalino L, Gillies RR, Shortell SM, Schmittiel JA, Bodenheimer T, Robinson JC, Rundall T, Oswald N, Schauffler H, Wang MC. External incentives, information technology, and organized processes to improve health care quality for patients with chronic diseases. *Journal of the American Medical Association*. 2003.
- [4] Aljedaani, Bakheet, Aakash Ahmad, Mansoorah Zahedi, and M. Ali Babar. "An Empirical Study on Developing Secure Mobile Health Apps: The Developers' Perspective." In 2020 27th Asia-Pacific Software Engineering Conference (APSEC), pp. 208-217. IEEE, 2020.
- [5] Ahmad, Aakash, Muhammad Waseem, Peng Liang, Mahdi Fahmideh, Mst Shamima Aktar, and Tommi Mikkonen. "Towards human-bot collaborative software architecting with chatgpt." In Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering, pp. 279-285. 2023.
- [6] Aljedaani, Bakheet, Aakash Ahmad, Mansoorah Zahedi, and M. Ali Babar. "Security awareness of end-users of mobile health applications: an empirical study." In *MobiQuitous 2020-17th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pp. 125-136. 2020.
- [7] Detmer DE, Munger BS, Lehmann CU. *Clinical informatics: principles, practice, and products*. Washington, DC: American Medical Informatics Association; 2019.
- [8] Garg R, Aggarwal A, Gupta N, Sharma A. Picture archiving and communication systems (PACS): A systematic review of published literature of the last decade and future directions. *J Med Syst*. 2018;42(2):25. doi: 10.1007/s10916-017-0916-4.
- [9] Johnson CD. Radiology Picture Archiving and Communication Systems (PACS): An Introduction. *J Digit Imaging*. 2018;31(3):283-291. doi: 10.1007/s10278-018-0073-0.
- [10] Bates DW, Gawande AA. Improving safety with information technology. *N Engl J Med*. 2003;348(25):2526-2534. doi: 10.1056/NEJMs020847.
- [11] Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to Electronic Health Record Adoption: a systematic literature review. *J Med Syst*. 2016;40(12):252. doi: 10.1007/s10916-016-0628-9.
- [12] Halamka, J.D.; Lippman, A.; Ekblaw, A. The Potential for Blockchain to Transform Electronic Health Records. *Harvard Bus. Rev*. 2017, 3, 2–5. Available online: <https://hbr.org/2017/03/the-potential-for-blockchain-to-transform-electronic-health-records> (March 20, 2020).
- [13] Benet J. IPFS - Towards a Permanent Web. In: *ACM SIGCOMM Computer Communication Review*. 2015;45(3):1-2.
- [14] Wang Z, Choo KKR. Medicalchain: A Secure and Lightweight Blockchain-Based Framework for Medical Data Sharing. *J Med Syst*. 2018;42(8):136. doi: 10.1007/s10916-018-1004-9.
- [15] Maldjian PD, Lee JH. Advances in PACS: Picture Archiving and Communication Systems. *Radiol Clin North Am*. 2009;47(6):1051-1064. doi: 10.1016/j.rcl.2009.07.009.
- [16] Zhang Y, Liu H, Guo Y. Research on privacy protection technology of medical image. In: 2017 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE; 2017. p. 36-39. doi: 10.1109/ICCCBDA.2017.8073701.
- [17] Hu F, Cai W, Feng D, Chen M. Challenges of medical big data analysis in developing intelligent healthcare. *J Healthc Eng*. 2018;2018:5061561. doi: 10.1155/2018/5061561.
- [18] Lin H, Huang H, Zhang R, Zhang S, Xie G. Integrated medical image management and analysis system using big data technologies. *IEEE J Biomed Health Inform*. 2017;21(1):143-154. doi: 10.1109/JBHL.2016.2633738.
- [19] Robson C. Privacy and security of electronic health records. In: Liu D, editor. *Medical Imaging Informatics*. Springer; 2020. p. 75-93. doi: 10.1007/978-3-030-16547-4_5.
- [20] Dou D, Wang X, Zhang X, Zeng X. Privacy-Preserving and Scalable Medical Data Sharing Through Blockchain. *IEEE Trans Serv Comput*. 2021;14(3):666-680. doi: 10.1109/TSC.2020.2974521.
- [21] Li J, Yu J, Zhang C, Cao J, Zhang L, Xia F. A Medical Image Retrieval Approach Based on Blockchain Technology. *IEEE Access*. 2019;7:98952-98960. doi: 10.1109/ACCESS.2019.2924085.
- [22] Yu S, Chen Q, Wu L, et al. Intelligent Medical Data Management Using Blockchain Technology. *J Med Syst*. 2018;42(7):130. doi: 10.1007/s10916-018-0990-y.
- [23] Menachemi N, Collum TH. Benefits and drawbacks of electronic health record systems. *Risk Manag Healthc Policy*. 2011;4:47-55. doi: 10.2147/RMHP.S12985.
- [24] Huang HK. *PACS and Imaging Informatics: Basic Principles and Applications*. 2nd ed. Wiley; 2015.
- [25] Zhou X, Liu B, Yu L, Wang T, Jiang M, Zhang Y. Medical Image Retrieval: A Multimodal Approach. *J Healthc Eng*. 2017;2017:9348252. doi: 10.1155/2017/9348252.

- [26] Zvárová J, Hřebíček J, Kolářová J. Security and privacy issues in electronic health records: A systematic literature review. *Health Informatics J.* 2020;26(1):30-45. doi: 10.1177/1460458219852984.
- [27] Dou D, Wang X, Zhang X, Zeng X. Privacy-Preserving and Scalable Medical Data Sharing Through Blockchain. *IEEE Trans Serv Comput.* 2021;14(3):666-680. doi: 10.1109/TSC.2020.2974521.
- [28] Benet J. IPFS - Content Addressed, Versioned, P2P File System. arXiv preprint arXiv:1407.3561. 2014.
- [29] Dagher GG, Mohler J, Milojkovic M, Marella PB. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sens Smart Cities.* 2017;97:1-8. doi: 10.1109/SESC.2017.7940157.
- [30] Meena M, Yadav AK. Blockchain based Secured Electronic Health Record Sharing System. In: 2020 International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE; 2020.
- [31] C. Manville, G. Cochrane, J. Cave, J. Millard, J. K. Pederson, R. K. Thaarup, A. Liebe, M. Wissner, R. Massink, B. Kotterink, Mapping smart cities in the eu (2014).

An Ensemble Learning Approach for Multi-Modal Medical Image Fusion using Deep Convolutional Neural Networks

Andino Maselena¹, Dr. D. Kavitha², Koudegai Ashok³, Dr. Mohammed Saleh Al Ansari⁴,
Nimmatai Satheesh⁵, Dr. R. Vijaya Kumar Reddy⁶

Institut Bakti Nusantara, Lampung, Indonesia¹

Associate Professor, Department of Information Technology, PVP Siddhartha Institute of Technology, Vijayawada²
Associate Professor, Vignana Bharathi Institute of Technology, Ghatkesar, Hyderabad³

Associate Professor, College of Engineering, Department of Chemical Engineering, University of Bahrain, Bahrain⁴
Department of Computer Applications, PSNA College of Engineering and Technology, Dindigul⁵

Associate professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India⁶

Abstract—Medical image fusion plays a vital role in enhancing the quality and accuracy of diagnostic procedures by integrating complementary information from multiple imaging modalities. In this study, we propose an ensemble learning approach for multi-modal medical image fusion utilizing deep convolutional neural networks (DCNNs) to predict brain tumour. The proposed method aims to exploit the inherent characteristics of different modalities and leverage the power of CNNs for improved fusion results. The Generative Adversarial Network (GAN) strengthens the input images. The ensemble learning framework comprises two main stages. Firstly, a set of DCNN models is trained independently on the respective input modalities, extracting high-level features that capture modality-specific information. Each DCNN model is fine-tuned to optimize its performance for fusion. Secondly, a fusion module is designed to aggregate the individual modality features and generate a fused image. The fusion module employs a weighted averaging technique to assign appropriate weights to the features based on their relevance and significance. The fused image obtained through this process exhibits enhanced spatial details and improved overall quality compared to the individual modalities. On a diversified dataset made up of multi-modal medical images, thorough tests are carried out to assess the efficacy of the suggested approach. The fusion images exhibit improved visual quality, enhanced feature representation, and better preservation of diagnostic information. The BRATS 2018 dataset, which contains Multi-Modal MRI images and patients' healthcare information were used. The proposed method also demonstrates robustness across different medical imaging modalities, highlighting its versatility and potential for widespread adoption in clinical practice.

Keywords—Deep convolutional neural networks; image fusion; generative adversarial network; ensemble learning

I. INTRODUCTION

In recent years, the field of medical imaging has witnessed tremendous advancements with the availability of multiple imaging modalities. Each modality provides unique information about anatomical structures, functional processes, or disease characteristics, making it crucial to extract comprehensive insights by combining data from multiple

modalities [1]. To effectively utilize the complementary information present in multi-modal medical images, researchers have turned to image fusion techniques. Image fusion aims to integrate data from different modalities into a unified representation, allowing for enhanced visualization, improved diagnostic accuracy, and better decision-making in clinical settings [2]. Several tasks related to computer vision, such as picture categorization, object identification, and categorization, have shown DCNN to be quite effective. In recent years, DCNNs have also gained significant attention in the domain of medical image analysis due to their ability to automatically learn complex features from large-scale data [3].

With the use of convolutional neural networks with deep layers, this work suggests a collaborative learning approach for fusing multimodal medical images. This method aims to overcome the limitations of traditional fusion methods by automatically learning the optimal fusion strategy from the data itself. The ensemble learning framework involves training multiple DCNNs, each specializing in capturing distinct features from different modalities [4]. These networks are designed to learn a shared representation that combines the information from each modality effectively. The ensemble is formed by aggregating the predictions of these networks, yielding a fused image that encapsulates the strengths of each modality [5]. By adopting an ensemble approach, this method leverages the diversity and complementary nature of the individual networks, resulting in a more robust and accurate fusion outcome. Additionally, the ensemble enables us to address uncertainties associated with the fusion process by providing a measure of confidence for the final fused image [6].

We carried out comprehensive experiments on a wide range of multi-modal healthcare imaging datasets in order to assess the performance of the suggested technique. The results demonstrate the superiority of ensemble learning approach over traditional fusion methods and even single DCNN-based fusion techniques. The fused images exhibit improved clarity, enhanced structural details, and better discrimination of abnormal regions, making them highly valuable for clinical

decision support and medical research. ensemble learning approach for multi-modal medical image fusion, employing deep convolutional neural networks, offers a promising solution for extracting comprehensive information from multi-modal medical images [7]. By effectively integrating the strengths of different imaging modalities, this method holds the potential to advance the field of medical imaging and facilitate more accurate and informed clinical diagnoses. The classification of medical images is crucial to both medical management and educational endeavors [8]. The traditional approach's performance has reached its apex, though. In addition, it takes a lot of effort and time to extract and select classification parameters when using them [9]. The DCNN is an innovative approach to machine learning that has proven beneficial for a variety of categorization issues. CNN excel at a number of picture categorization tasks, producing the best results. However, medical image collections are difficult to compile since classifying them calls for an exceptionally high degree of professional competency [10].

Deep Convolutional Neural Networks (DCNNs) have also been employed for medical image fusion, offering several applications in healthcare. DCNNs can fuse low-resolution medical images with high-resolution images to generate enhanced, high-resolution images. This technique can be particularly beneficial in medical imaging, where higher resolution can provide better visualization of fine details, aiding in accurate diagnosis and treatment planning. DCNN-based fusion methods can fuse multiple images to generate a fused image with improved segmentation accuracy. By integrating information from different imaging modalities or perspectives, the fused image can provide more accurate and reliable boundaries and regions of interest for subsequent analysis and treatment planning [11].

DCNNs can be utilized for medical image registration, which involves aligning images from different modalities or time points. By fusing information from multiple images, DCNN-based methods can improve the accuracy and robustness of image registration, allowing for more precise analysis, monitoring, and treatment planning. It can fuse images from different sources to synthesize new images with desired characteristics or properties [12]. For example, fusing images from different imaging modalities can create a synthesized image that combines the strengths of each modality, providing comprehensive information for clinical analysis and decision-making. DCNNs can be employed for restoring medical images that are corrupted by noise, artifacts, or other degradations [13]. By fusing multiple degraded images, DCNN-based methods can effectively de-noise and enhance the image quality, enabling better visualization and interpretation of medical conditions. These applications demonstrate the versatility and effectiveness of DCNN-based medical image fusion techniques in improving image quality, accuracy, and clinical decision-making in various healthcare scenarios [14].

The key Contributions of this Research work is:

- The ensemble learning framework involves training multiple DCNN models independently on respective

input modalities, capturing modality-specific information.

- A fusion module is designed to combine the extracted features from individual modalities, employing a weighted averaging technique to assign relevant and significant weights.
- The fused image obtained through this process exhibits improved spatial details, enhanced feature representation, and better preservation of diagnostic information.
- Thorough testing on a diverse dataset confirms the efficacy, visual quality, and robustness of the proposed method, showcasing its potential for broad adoption in clinical practice.

The manuscript of the approached paper is organized as follows: In Section II, some related works are reviewed. In Section III, Information regarding the problem statement is provided. In Section IV, the proposed Multi-Modal Image Fusion is covered in detail. In Section V, experiment results are provided, and discussed in Section VI with an extensive evaluation of the proposed approach to current best practices is made. In Section VII, the conclusion of the paper is provided.

II. RELATED WORKS

Maqsood et al. [15] suggested a multimodal fusion of images approach is based on limited representation and two-scale picture segmentation. The original heterogeneous images are initially subjected to contrast enrichment processing in the proposed system, which improves the brightness distribution for better visualization. The edge data gathered from intensity extended images is extracted using a spatial gradient-based edge detection method. The fundamental and detail layers are separated from the improved multiple mediums images at this point. Utilizing SSGSM, the final detailed layer is extracted. Finally, the fused image is produced utilizing an improved judgement maps and fusion scheme. By conducting both quantitative and qualitative evaluations, the experimental results demonstrate that the recommended multimodal picture fusion strategy outperforms several previous methods. However, it could happen for certain data from the initial images to be destroyed or distorted during the fusion process. The fusion mechanism may prioritize some qualities or aspects while ignoring others, resulting in the loss of crucial information or subtle traits.

Dinh et al.[16] proposed that the following are the key phases in the unique strategy that was presented to address the aforementioned shortcomings. In order to acquire the basic and detail elements, the three-scale deconstruction (TSD) approach is initially presented. Second, the output picture is fused using a rule based on the nearby energy function and the Kirsch compass operator, which aids in the retention of critical information. Thirdly, to fuse base layers with the best characteristics and produce a high-quality picture, the Marine Processors Algorithm (MPA) is used. This work compared the effectiveness of the suggested technique using six photograph quality criteria and five cutting-edge medical image fusion

algorithms. Experiments revealed that the proposed method significantly increased the level of quality of the fusion picture and preserved edge information. The particular fusion algorithm used has a significant impact on the effectiveness of multidimensional picture fusion. Additionally, there doesn't exist a one-size-fits-all solution, and different techniques may yield different fusion results. The effectiveness and level of quality of the combined image can be considerably impacted by the algorithm choice.

Diwakar et al. [17] proposed a novel shearlet region multiple modalities image fusion method. The recommended technique uses Non-Subsampled Shearlet Transformation (NSST) to separate input pictures into low- and high-frequency parts. The localized extrema (LE) method is a unique technique used to separate and merge the fundamental layer and details layers. The co-occurring filter (CoF) is then used to combine the foundation layer and detail layer in harmonics with smaller elements. A high-frequency component is integrated using a sum modulated Laplacian (SML) as a component of an edge-preserving technique to image fusion. On the Multi-modal healthcare picture collection, experimental findings and contrasting assessment are performed using both recommended and modern methodologies. The recommended strategy beats cutting-edge fusion techniques in terms of blade retention in both objective and subjective assessment requirements, according to test findings and assessments. Numerous multidimensional merging of images algorithms is computationally demanding, requiring a significant amount of time and computing capacity. This could be a drawback in situations or real-time applications that need for quick fusion.

Stimpel et al. [18] demonstrated the globally linear guided filter for general medical image processing when coupled with a learning guiding map. The guided filter is the only element processing the output images, and its direction map may be trained to do the task optimally from beginning to end. The demising and graphic high-resolution tests are the two most often used activities when using this method to measure performance. The evaluation is based on cross-modal data sets that are paired. Modern methods are coupled with the provided procedure to achieve both goals. This can also show that the input image's information is basically unaltered after treatment, in contrast to conventional deep neural network approaches. The suggested pipeline also offers greater resilience against adversarial attacks and deteriorated input. Image fusion requires accurate registration of images from different modalities to align corresponding anatomical or functional structures. However, image registration can be challenging due to differences in acquisition protocols, patient motion, and anatomical variations. Registration errors can lead to misalignment and distortions in the fused image, affecting the accuracy of subsequent analysis.

Asha et al. [19] suggested a chaotic grey wolf optimization algorithm-based balanced blending of high-energy sub-bands of the Non-Subsampled Shearlet Transform (NSST) domain. The raw images are first dissected into their many scales and multi-directional components using the NSST. The modest number of pathways were combined according to a simple maximum rule in order to sustain the energy of an individual.

In order to combine images of various frequencies and minimize the difference between the resultant image and the starting point pictures while retaining the textural characteristics of the input images, a collection of automatically adjusted high-frequency images is used. The major goals of the entire procedure are to maintain the energy of a low-frequency region while transferring textural details from the source images to the fused image. In order to construct the fused picture, the inverse NSST of the combining minimal and high-energy bands is used. Eight distinct illness datasets from Brain Atlas are used in the trials. More than 100 picture pairings are used to evaluate the efficacy of the suggested strategy using both objective and subjective quality evaluation. Due to the lack of contemporaneous collection of several modalities or the difficulty in gathering ground truth annotation for fusion quality, obtaining grounding truth for multipurpose fusion in medical imaging is problematic. Due to this, evaluating and comparing fusion procedures quantitatively is more difficult and frequently relies on opinions or substitute measurements.

Li et al. [20] To address the issue of poor contrast detail, a powerful image fusion technique employing numerous prominent features and a guided image filter was presented. The input photos were first divided into a number of calming and thorough images that had different scales before being subjected to the directed picture filter. Second, two different algorithms are used to extract important characteristics from the broken-down dependent upon visuals alongside the complete images in order to develop the combination rules. These two algorithms are the spectral residual (SR) technique for the mainframe gathering and the graph-based visually prominence model for a gradient saliency information extraction. The decomposition factors are combined using a process known as generalized intensity-hue-saturation (GIHS). The fused image is then reconstructed from the combined smoother and detailed images. The experimental findings show that, in the fields of MRI-PET and MRI-SPECT fusion, the proposed algorithm can outperform previous fusion approaches. The acceptability and use of fusion procedures in clinical practice, where openness and comprehensibility are vital, may be hampered by this lack of comprehension. The availability of information for the various modalities in multipurpose medical imaging may not be equal, meaning that one modality may contain greater numbers of specimens than the others. The fusion process may be impacted by this modality imbalance, which might result in biased fusion findings or a restricted representation of less common modalities.

Dai et al. [21] suggested that transformers have enormous promise for multimodal medical picture categorization. The proposed approach is based on the successful extraction of the link among sequences by the transformer. However, due to the small dimensions of medical information sets for pictures and the lack of sufficient data to establish the connection between low-level semantic variables, the precision of pure transformation systems based on ViT and DeiT is not good in versatile classification of medical images. TransMed is therefore suggested as a way to collect both cross-modality high-level information and low-level characteristics.

TransMed combines the benefits of both CNN and Transformer. TransMed converts the multimodal pictures into sequences, delivers them to CNN for processing, and then use transformers to discover the connections between each sequence and provide predictions. TransMed beats the current multipurpose fusion approaches when it comes to of parameters, operating speed, and accuracy because the transformer successfully models the global aspects of multifaceted pictures. Finding the best fusion approach, though, is a challenging task. Different fusion methods, each with various advantages and disadvantages, may be used, including pixel-level, decision-level fusion and feature-level. For a certain application or modality combination, choosing the best fusion approach necessitates extensive thought and skill.

III. PROBLEM STATEMENT

Multi-modal medical imaging provides valuable complementary information for accurate diagnosis, treatment planning, and monitoring of various diseases. The issue of successfully integrating and fusing data from many imaging modalities is still difficult. Traditional fusion methods' dependence on ad hoc extraction of features and fusion techniques that commonly use handmade feature extraction that approximate complicated interactions between paradigms may restrict the quality of the merged image. Furthermore, the actual applicability of these approaches in clinical contexts is hampered by their lack of stability and interpretability. CNN have proven to perform exceptionally well in a variety

of computer vision applications, including the processing of medical images. CNNs have not yet been extensively used in multi-modal medical picture fusion, nevertheless. The research gap in the mentioned existing works lies in the need for more comprehensive and adaptable multimodal image fusion techniques that can simultaneously address various aspects of quality enhancement, edge preservation, and overall visual fidelity. The proposed DCNN aims to overcome the limitations of traditional fusion methods and single CNN-based approaches by effectively capturing the complementary information present in multiple modalities and improving the fusion quality. The ensemble learning framework is expected to leverage the diversity and strengths of individual networks to enhance the accuracy, robustness, and interpretability of the fusion process [22].

IV. PROPOSED ENSEMBLE LEARNING APPROACH

The suggested method entails enhancing the supplied image. Then DCNN are used to accomplish Medical Image Fusion. The performance is then assessed. The suggested method for Multi-Modal Medical Image Fusion using DCNN is shown in Fig. 1. The input photographs are first preprocessed by converting them to a standard scale and using the proper transformations to improve image details. Then, using a sizable dataset of aligned multi-modal pictures and a fusion-specific loss function, a CNN architecture is created, consisting of shared and modality-specific convolutional layers. From each modality, high-level feature maps are retrieved using the trained CNN.

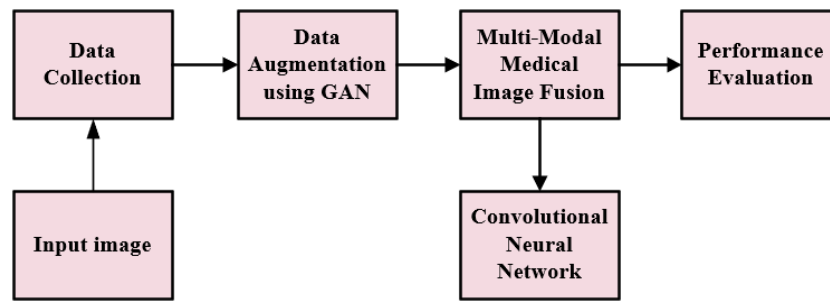


Fig. 1. Proposed approach for multi-modal medical image fusion.

A. Data Collection

The BRATS 2018 dataset, which contains Multi-Modal MRI images and patients' healthcare information with distinct heterogeneous histologic sub-regions, different levels of aggression, and variable prognosis, was used for training and testing in this work. These clinical multi-modal MR images have been generated using a range of magnetic field intensities and scanners [23]. Table I describe the dataset for Training and Validation.

TABLE I. THE COLLECTED DATASETS

	Training data	Testing data
Unhealthy	350	350
Healthy	250	250
Overall data	600	600

B. Data Augmentation using GAN

Data analysis, enhancement, combination, and rescaling are all part of preprocessing. The acquired source photographs are transformed to RGB images before augmentation. The improvement method is used to build more powerful simpler models that are impervious to some sorts of picture manipulation, following which the image's quality is altered to improve the information's integrity and degree of variability. The layering placement of the original photos is important for the concatenation. The first phase is the R channel of an MRI, and the second layer is the R channel of a PET scan (positron emission tomography). In the second layer, the B(Blue) channels of MRI and the B channels of PET are placed after the G(Green) channels of MRI and the G channel of PET, respectively. The pictures that offer practical details must be maintained below the photos that offer structural details in all pathways, it must be highlighted [24].

Fig. 2 the generator creates a picture according to the parameters that are collected from the image, based on the number of channels supplied in the layers for input and output. The modelled output picture of the suggested method contains three channels and six input channels. The generator automatically fits the obtained parameters into the three stated channels during training. The stacked order is set to RR1GG1BB1 and the training data is compressed to prevent the color space from being disrupted. The RGB components of the first source picture are R, G, and B, while the RGB components of the additional source image are R1G1B1. Random switching and unpredictability are employed for data augmentation. With random flipping, there is a 50% chance that the picture will be turned. The picture is accompanied with random noise, which is Gaussian in nature with an average value of 0 and a variation of 0.1. This method can be used to learn the aggregate breakdown of single-modality imaging information as well as for recording the broad distribution of imaging data from several modalities. The primary producer can learn to produce many modalities at once since different modalities' information collected from a single ROI share identical information with unique appearance patterns. Such a generator can be used to complete missing modes of operation or supplement data [25].

As an estimate $K_{data}(u)$, GAN aims to learn an estimate of probabilities, $k_G(u)$, from the actual distribution. $u = G(v)$, the sample, where the noise variable is called v . It resolves the issue by simultaneously instructing the generator N and a discriminator D to create a process that is adversarial. By sampling noise, G produces samples from latent space. Whether the sample comes from $K_G(u)$ or $K_{data}(u)$ is determined by D . G samples eventually approach genuine or real samples through the continuous unfavorable effect. The definition of the optimization formula D is represented in Eq. (1) [26].

$$D^* = \operatorname{argmin} \operatorname{Div}_D(k_D(u), k_{data}(u)) \quad (1)$$

Where $\operatorname{Div} (*)$ indicates the divergence among the two distributions. N may be used to compute the divergence and generate the following objective function as represented in Eq. (2)

$$N^* = \operatorname{argmax} V_F(G, D) \quad (2)$$

Where,

$$V(G, D) = J_{X \sim K_{data}} [\log D(x) + j_{X \sim k_G} [\log(1 - D(u))]] \quad (3)$$

Hence, the Eq. (1) is transformed as

$$G^* = \operatorname{argmin}_N \max_D V(N, D) \quad (4)$$

In contrast to a traditional GAN, which consists of a single generator and a discriminator, pix2pixHD uses an auxiliary

producer and a primary generator to output pictures at two distinct resolutions, which are $3 \times 448 \times 448$ and $3 \times 224 \times 224$ in this instance. Therefore, two entirely convolutional network-based discrimination named D_p and D_q are in charge of the two solutions.

C. Multi-Modal Fusion-CNN

Patch incorporation, class insertion, position integration, class token and patch token are the five insertions and tokens that are present in the input layer. While class anchoring is an adaptable vector, patch anchoring represents each patches' input from CNN. Using position embedded data and patched embedded data; this technique preserves the geographical and geographical data of a patch by encoding it into patch tokens. Class signaling and class anchoring are equal since category anchoring does not provide patch embedding. The Eq. (5) and (6) represents the input is u , the adaptable vector is V_a , the location embedding is u_{pa} , the patch tokens are u_{pq} , and the class token is u_{de} .

$$u_{pq} = \operatorname{Conv}(u) + u_{pa} \quad (5)$$

$$u_{de} = V^a \quad (6)$$

The type token connects to the patched tokens preceding the converters' input layer, goes via the conversion layer, and is subsequently generated from the fully connected layer in order to foresee the class. The core of the arrangement is an image power source, which receives images from various input modalities and generates a task-optimal unified depiction of the required guiding map. In extracting the most important information directly from data, convolutional neural networks (CNN) have demonstrated significant success. A CNN is applied to build the guiding map as a result. De-noising and picture super resolution are the two tasks we focus on. The guide maps for both are generated using tested network designs for the sake of repeatability. The necessity to handle numerous input photos led to the sole adjustments. The inclusion of more guiding photos would logically be conceivable and is only constrained by availability and processing capacity. In order to determine how the selected network design affects the guided filtering process, employs using two separate networks for super resolution [27].

Fig. 3 represents the CNN architecture where a layer of neurons is fully linked, every neuron in that layer is also connected to every neuron in the layer underneath it. The value should indicate the degree to which of the connection between the neuron that is j^{th} in this particular stratum and the k^{th} neurons in the preceding layer ∂_{0lk} . Let b_{1j} be the bias of the j^{th} neuron in the current layer. The result of the layer's j^{th} neuron is given by Eq. (7).

$$y_{0j} = \sum_k \partial_{1lk} x_{0k} + b_{1j} \quad (7)$$

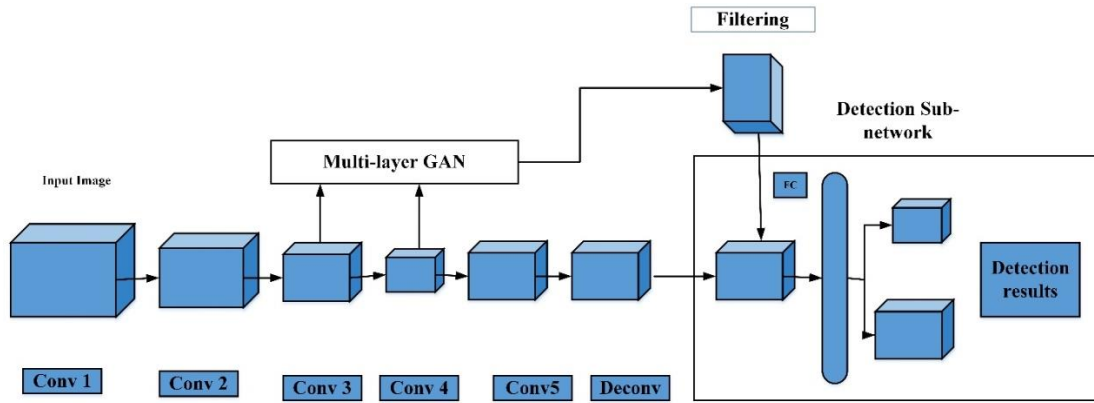


Fig. 2. GAN in data augmentation.

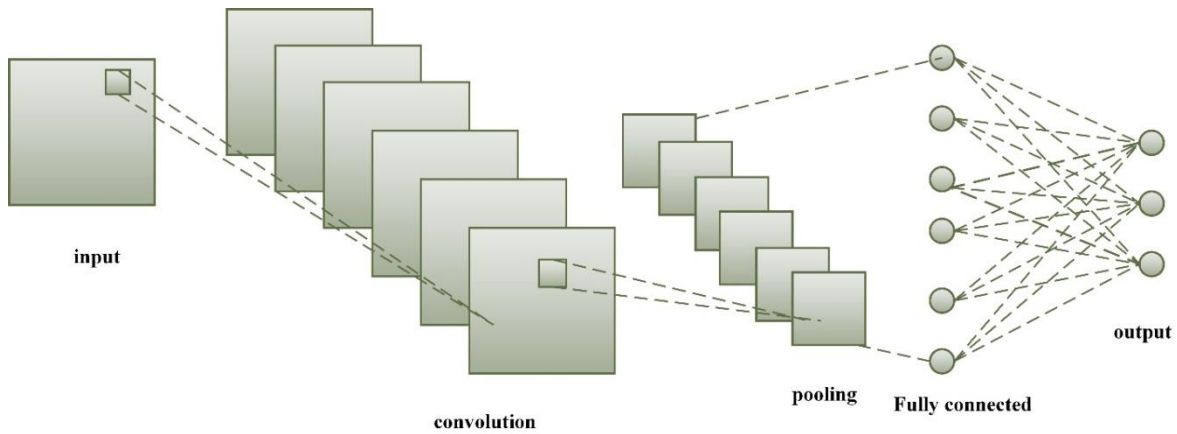


Fig. 3. CNN architecture.

The convolutional layer's neurons that are frequently utilized to produce a kernel or filtration have the same biases and values. If the dimensions of the filtering are set to $n \times n$, every neuron in the corresponding layer will be linked to a $n \times n$ area of the neurons that are in the layer above it. In line with this, the $(j, k)^{\text{th}}$ neuron's outputs will be in Eq. (8)

$$y_{1,j,k} = \sum_{p=0}^{n-1} \sum_{q=0}^{n-1} \omega_{0,p,q} x_{0,j+p,k+q} + c \quad (8)$$

Examples of regularly used activation functions include Tanh, Sigmoid, and the inverted linear unit, which is now the de facto recommendation for contemporary neural networks. Convolutional or completely linked layers are typically followed by activation layers to provide elementwise non-linear behavior. By using the activation function, is defined in Eq. (9)

$$f_0(x) = \max(x, 0) \quad (9)$$

The downwards sampling method for every sub-area in the pooling layer provides the dimension of a single neuron in the present one by dividing the neurons of the layer preceding it into an array of not overlapping rectangles. Maximum pooling and average-pooling, the two most popular pooling procedures, offer the subarea's maximum value and average value, respectively. A convolutional neural network usually sets up a sequence of convolutional (Conv)-ReLU layers, before adding the pooling layers (Pool), and continues doing this till the picture gets spatially combined to a compact size.

At certain points, it is usual to switch to fully-connected layers (FC). Three different parts make up each neuron: the receiving domain, a modulation domain, and a pulse-generating domain. Feedback is created by the connections between many neurons. An external input signal is first received in the receiving domain, after which it is amplified in the modulating domain and the final output pulse is produced in the pulse generating domain. The fundamental procedure is as follows: signals from the feedback channel domain and the link domain are received in the receiving domain, and they travel through Channels L and F into the modulation domain.

The necessary feature pixels in the layer of convolution are added to each image's output pixel after synchronizing the characteristics from the source pictures. Add every value of a pixel together, and then divide the result by the total number of pixels in the description. The feature map has been added to the computed values, causing the improvement to be applied to the whole image. The characteristics map has a slot for each computed value. All of the traits are therefore processed, and several feature maps are produced. The Eq. (10) to obtain the convolutional layer is the following,

$$v_{xyz} = \sum_{E=0}^{E-1} \sum_{F=0}^{F-1} \sum_{H=0}^{G-1} s_{x+g,y+g,E}^{(l-1)} e_{eghe} + f_{pqr} \quad (10)$$

Where f_{pqr} is generally set to which is not contingent on the image's component position. $E^{e}eghe$ as an identical value of weight. Since it recovers the distinguishing properties of the

image using various convolution kernel sizes, the layer of convolution is a critical part of CNN. The layers of inversion can be continuously applied to the input photos to create a set of feature maps. K_i may then be created by using the characteristic map of the i -th layer in CNN as it is represented in Eq. (11)

$$K_i = \rho(K_{i-1}V_i + H_i) \quad (11)$$

Where K_i is the current networks layer's mapping of features, D_{i-1} is the previous layer's convolution feature. V_i is the i -th layer weight, k_i is the i -th layer offset vector, and $\rho(\cdot)$ represents the rectified function. Layer pooling's goal is to reduce the total amount of space, that can cut processing costs and effectively reduce the danger of over-fitting. The resultant characteristic on the i th localized responsive field is determined in Eq. (12) in the k -th layer of pooling.

$$u_i^k = \text{down}(u_i^{k-1}, s) \quad (12)$$

where $\text{down}(\cdot)$ indicates the function for down-sampling, u_i^{k-1} is the feature vector in the previous layer, and r is the pooling size. Following the pooling and convolutional layers, there may be one or more fully-connected (FC) layers, which use the collected features for picture categorization. It classifies the input brain images into healthy and unhealthy.

D. Multi-Modal Fusion

In order to process sources of any size, the conversion phase is employed throughout the picture checking and fusion procedure on the totally connected layer. Using the same kernel size, the entire connected layer is split into two comparable convolutional layers. The network may then combine pictures of any size, X and Y , to produce a dense prediction map, I . Each prediction I_s on the map is represented by a vector with two dimensions with values between 0 and 1. If one dimension of a prediction is bigger than the other, it is normalized to 1 while the other dimension is set to 0, making the weights given to related image blocks easier. With an

outcome aspect value of 1, this ensures that the weight of every image block is decreased. Two near forecasts in S have overlapping areas in their corresponding picture blocks. The weights of the photos in these overlapped portions are added to determine the mean value of the adjacent picture blocks. The network may be given pictures of any size, both X and Y , using this technique, and a weight map W of the same size is generated. This guarantees a weight reduction for each picture block with an output aspect value of 1.

E. Fusion Rules

In order to attain better look, richer details, and spectacular fusion impacts, this study suggests novel fusion principles and the average weighted fusion operations in accordance with area peculiarities. The fusion guidelines and commands are as follows:

Stage 1: It determines the energy R_u^o and R_v^o of matching localized areas in each breakdown layer o of source images x and y , accordingly, using the contrast pyramid deconstruction:

$$R_u^o(a, b) = \sum_x \sum_y S_u^o(a + u, b + m)^2 \quad (13)$$

$$R_v^o(a, b) = \sum_x \sum_y S_v^o(a + u, b + m)^2 \quad (14)$$

Where Equations (13) and (14) the regional area power $R_u^o(a, b)$ on the o^{th} layer of difference is centered at (a, b) . structure, where u and v stand for the size of the region in question, and represents the image of the contrast between the structuring fourth layer.

Stage 2: Determine how similar the respective local areas in two source photos are to one another.

Stage 3: Decide who the fusion operators.

As a consequence, the strategy selects the center pixel based on energy variations when the degree of similarity is below the threshold of significance and employs the weighted fusion operator when it is equal to or above.

Algorithm 1: Multi-Modal Medical Image Fusion using Deep Convolutional Neural Networks

Input: Medical Images

Output: fusion result

The two source images and the initial fused one are given

Train the input images v_i in the system, where $i = 1$ to n

Data Augmentation of images

Let $U(i)$ be the input images from the dataset

for every U_i

$$V_v(i) = V(i) - N$$

// using GAN

// V denotes unwanted noise

Segmentation of images

Initialize the starting point of the highlighted portion

if (image detected)

Gather the subset

Identify the highlights in the hyperspectral image using Eq. (7)

Else

Repeat until the stopping condition is reached

// until the image is identified

End if

Return

Image Fusion using CNN

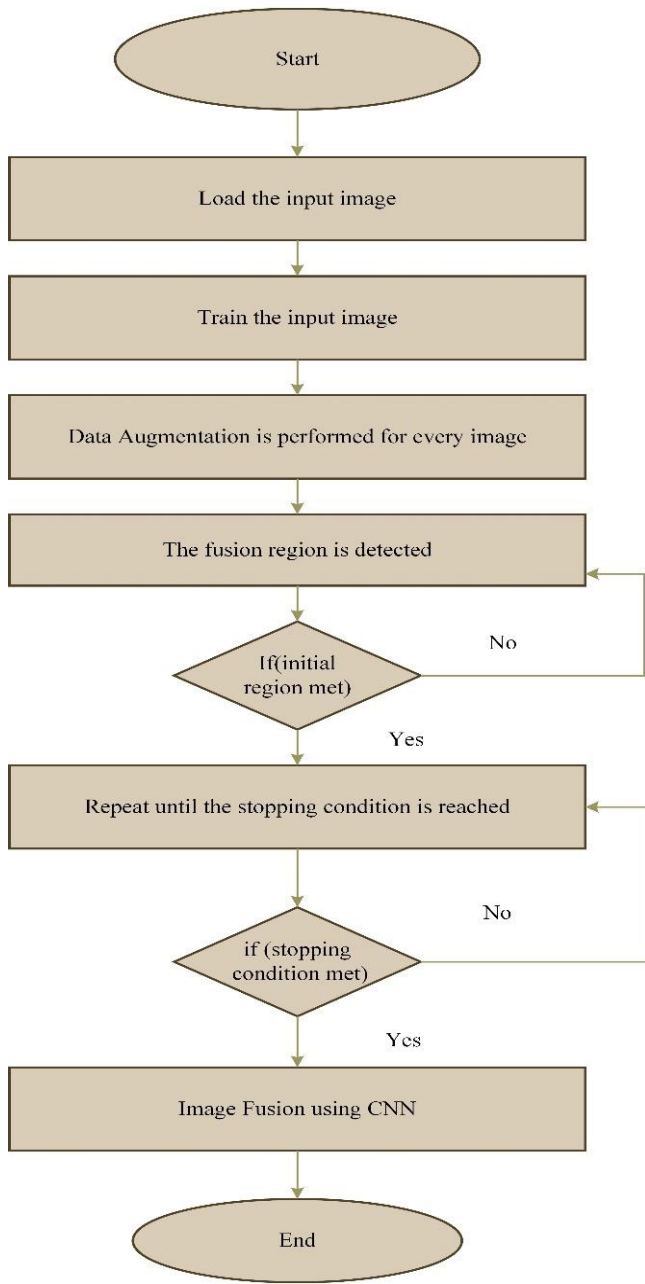


Fig. 4. Flow chart of the proposed system.

Fig. 4 represents the Ensemble Learning Approach for Multi-Modal Medical Image Fusion using Deep Convolutional Neural Networks.

V. RESULTS

The recommended method has been evaluated using datasets and executed in MATLAB software on the Windows 10 platform. In order to solve this issue, deep CNNs are utilized in the article to extract high-level characteristics from the data modalities and NMF is employed to discover the fused image's underlying structure. The use of deep CNNs, which have demonstrated extraordinary capacity in understanding intricate patterns and characteristics from pictures, is a key benefit of the suggested technique. The

model can accurately capture and reflect the unique qualities of each modality by utilizing the power of deep learning, thereby enabling a more thorough synthesis of information. In order to evaluate the effectiveness of their strategy, the authors additionally offer quantitative assessment criteria including precision, recall, precision, accuracy, F-score, specificity, and sensitivity. The suggested approach's robustness and dependability are highlighted by the excellent scores in these criteria that were attained. Overall, multi-modal image fusion using deep CNNs and NMF makes a significant addition to the discipline. The suggested approach successfully combines deep learning for feature extraction with NMF to train the fused representation, producing better fusion results. The results of the investigation and analyses show how this approach can be used for a range of applications, including mapping, and imaging in medicine. The use of multipurpose image fusion techniques in the health care imaging field is essential for better medical diagnosis and therapy. The research study suggests a unique method for fusing multimodal medical images that incorporates deep convolutional neural networks (CNNs).

A. Accuracy

The model's total Accuracy shows how well it performs across all classifications. In overall, it is the idea that every circumstance can be forecast with accuracy. Eq. (15) represents the Accuracy:

$$A = \frac{T_{pos} + T_{neg}}{T_{pos} + T_{neg} + F_{pos} + F_{neg}} \quad (15)$$

B. Precision

Precision is calculated as the total amount of positive predictions multiplied by the number of correct positive estimations. It measures how many accurately merged multi-modal medical pictures there are. Eq. (16), which is used to compute the accuracy

$$P = \frac{T_{pos}}{T_{pos} + F_{pos}} \quad (16)$$

C. Recall

The ratio of correct positive forecasts to true positives and false negatives is known as recall. It displays the proportion of correctly predicted events and picture fusion across different modes. The recall is represented by Eq. (17),

$$R = \frac{T_{pos}}{T_{pos} + F_{neg}} \quad (17)$$

D. F1-Score

Precision and recall are combined in the F1-Score calculation. The F1-Score as shown in Eq. (18) is created using precision and recall.

$$F = \frac{2 \times \text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}} \quad (18)$$

E. Sensitivity

It is a measure of the proportion of correctly foretold true positives. Eq. (19) is used to calculate sensitivity as,

$$\text{Sensitivity} = \frac{T_{pos}}{T_{pos} + T_{neg}} \quad (19)$$

F. Specificity

The degree gauges identify precisely the true negatives. Eq. (20) is used to calculate the specificity value as,

$$Specificity = \frac{T_{neg}}{F_{pos} + T_{neg}} \quad (20)$$

TABLE II. COMPARISON OF ACCURACY

Classifier	Accuracy
CNN [10]	86.8
RNN [11]	97.9
KNN [14]	98.2
AlexNet[16]	98.5
DCNN	99.6

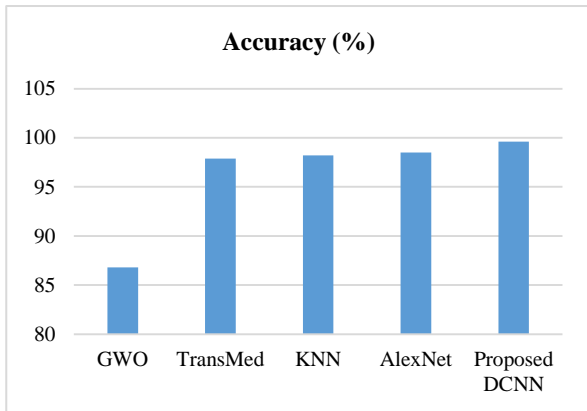


Fig. 5. Comparison of accuracy.

When compared to the current techniques, the suggested technique DCNN obtains a greater level of accuracy. The contrast of efficiency between DCNN and other approaches is shown in Table II and diagrammed in Fig. 5.

TABLE III. COMPARISON OF PRECISION AND RECALL

Methods	Precision (%)	Recall (%)
KNN	89.5	89.1
CNN	96.9	98.5
GWO	97.9	95
DCNN	99.9	99

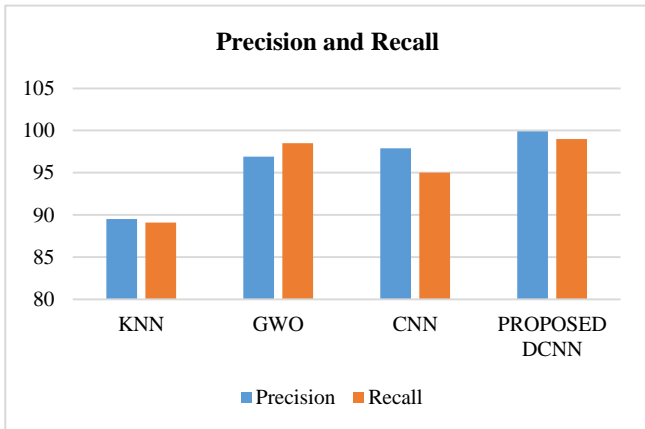


Fig. 6. Comparison of precision and recall.

Table III demonstrates that the proposed technique of combined DCNN achieves higher precision and recall of 99.9% and 99% when compared to the existing methods. The advanced DCNN gives better accuracy than the performance evaluated. Here, the achieved accuracy level is 99 using the DCNN model. Fig. 6 illustrates the precision and recall between DCNN and other methods. The model's balanced and trustworthy performance is further supported by the F-score which takes precision and recall into account. These findings support the suggested model's exceptional qualities, including precision, recall, precision, F-score, sensitivity, and specificity, which make it a trustworthy and efficient option for the task at issue.

TABLE IV. SENSITIVITY AND SPECIFICITY FOR PROPOSED METHOD

Proposed Model	
Sensitivity	98.14
Specificity	96.68

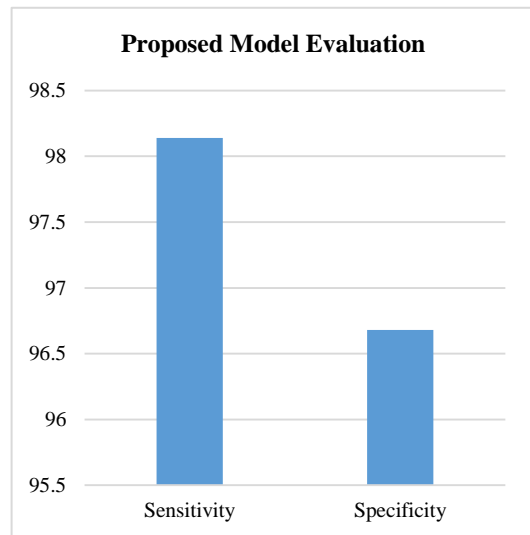


Fig. 7. Comparison of specificity and sensitivity.

Fig. 7 and Table IV represents the model's specificity score of 98.14% shows that there is little chance of making false positives for negative predictions, while its sensitivity score of 96.68% emphasizes how well the model can recognize positive circumstances.

The accuracy of the convolutional neural network used for both the training and testing stages is 99.4% and 97.5%, respectively, according to Table V. When DCNN is utilized, the accuracy of the testing and training processes increases to 99.9% and 99.4%, respectively. Fig. 8 shows an evaluation of performance.

TABLE V. PERFORMANCE EVALUATION

	CNN	ABO-CNN
Training	98.1	99.9
Testing	97.5	99.4

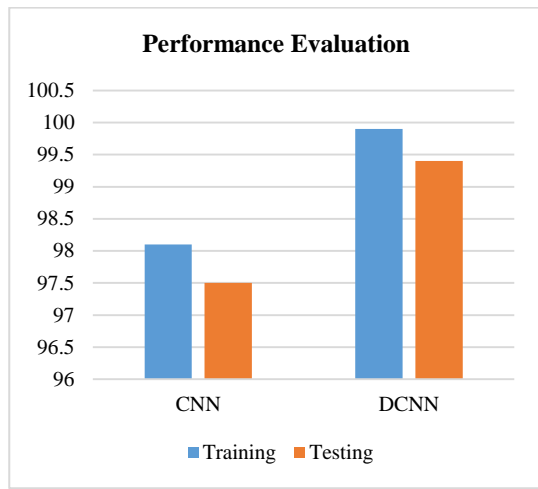


Fig. 8. Performance evaluation.

Table VI and Fig. 9 presents a comparison of medical image fusion techniques based on three evaluation metrics: Gradient-based quality, Information ratio, and Mutual information. Each metric is accompanied by corresponding percentages representing the performance of the techniques in relation to that metric. The Gradient-based quality metric is evaluated at 89%, 45.5%, and 67.7% for RMSE, indicating the percentage of quality achieved by the fusion techniques in terms of gradient-based measures. Similarly, the PSNR metric indicates a performance of 54%, 40%, and 79% for the techniques, representing the Peak Signal-to-Noise Ratio achieved by the fusion results. Lastly, the ASR metric is reported at 45%, 39.5%, and 59%, representing the Accuracy Success Rate of the fusion techniques. This table allows for a comparative analysis of different medical image fusion methods based on multiple evaluation metrics, providing insights into their respective performance levels across various quality measures.

TABLE VI. MEDICAL IMAGE FUSION COMPARISON

	Gradient-based quality	Information ratio	Mutual information
RMSE	89%	45.5%	67.7%
PSNR	54%	40%	79%
ASR	45%	39.5%	59%

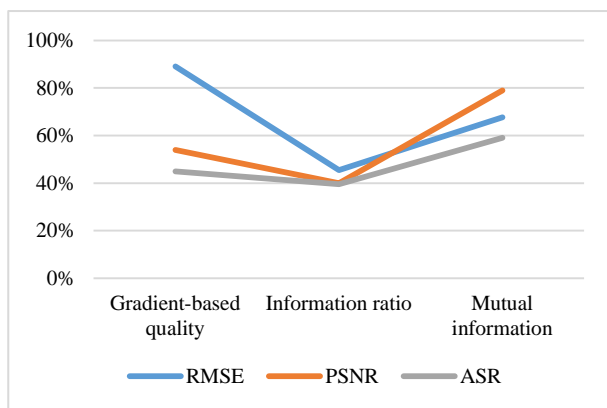


Fig. 9. Medical image fusion comparison.

TABLE VII. COMPARISON OF PROCESSING TIME

Methods	Processing Time
KNN	11.05
CNN	14.58
GWO	12.86
DCNN	6.15

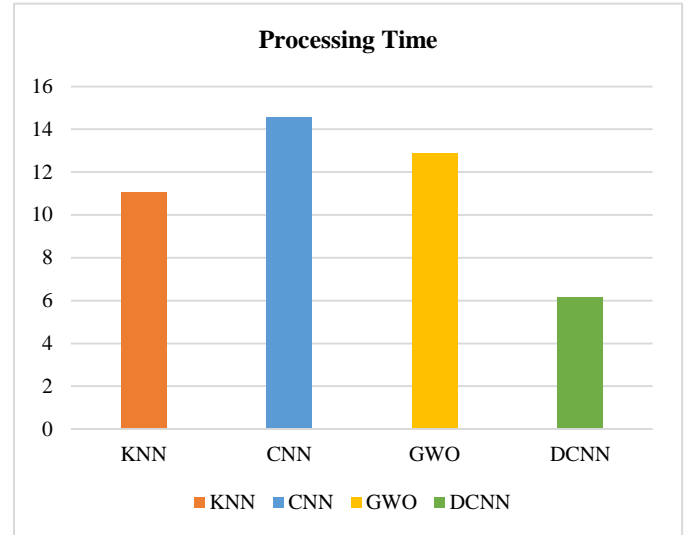


Fig. 10. Evaluation Comparison of processing time.

Table VII and Fig. 10 presents a comparison of processing times for different methods, namely KNN, CNN, GWO, and DCNN. The Processing Time column indicates the time taken by each method for a specific task or process. From the table, it can be observed that KNN takes 11.05 units of time, CNN takes 14.58 units, GWO takes 12.86 units, and DCNN takes 6.15 units. These values reflect the computational efficiency or speed of each method, with a lower processing time indicating faster execution. The table provides insights into the relative performance of these methods in terms of processing time, which can be valuable for selecting an appropriate method based on time constraints or efficiency requirements.

G. ROC Curve

Fig. 11 represents the ROC Curve of the proposed system. The proposed DCNN has the higher rate when compared to the existing methods. The ROC curve is a graphical representation of the performance of a binary classification system as its discrimination threshold is varied. However, the ROC curve is not directly applicable to evaluate multi-modal image fusion, as it is typically used for evaluating classification models.

H. Accuracy and Loss for Training and Validation

Fig. 12 represents the accuracy of a multi-modal image fusion model refers to how well it can effectively integrate and preserve relevant information from the input images while suppressing noise, artifacts, and inconsistencies.

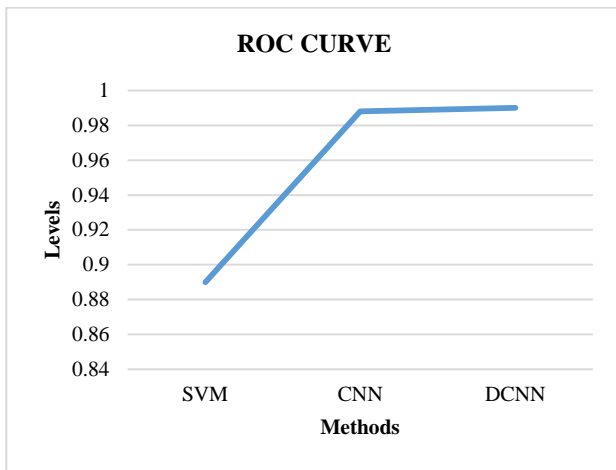


Fig. 11. ROC curve.

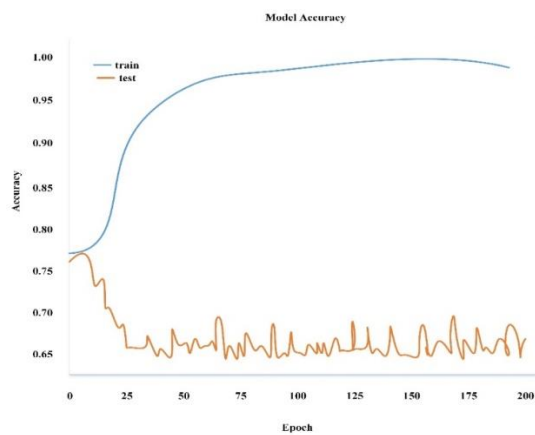


Fig. 12. Model accuracy for training and validation.

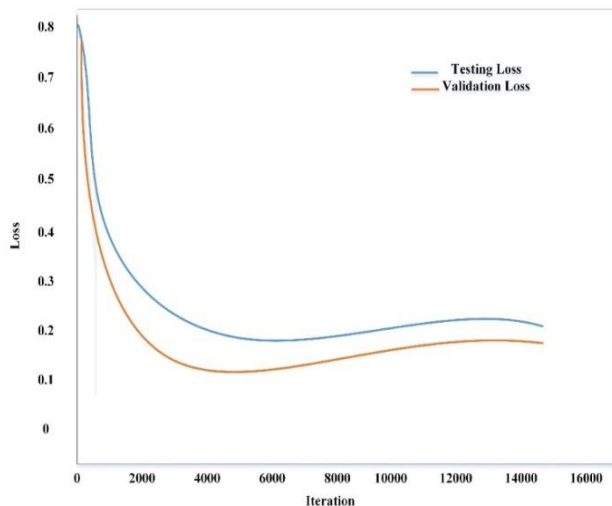


Fig. 13. Model loss for training and validation.

VI. DISCUSSION

Existing techniques frequently concentrate on certain aspects of image fusion, such as feature extraction, detail preservation, or computational effectiveness, but there are no complete solutions that handle all elements of quality, such as contrast augmentation, edge preservation, and overall visual fidelity. The performance of fusion approaches across diverse modalities, clinical applications, and data quantities cannot be fully evaluated due to the lack of defined assessment parameters. It is still difficult to get timely collections of multi-modal data and trustworthy annotations, hence the problem of ground truth annotation for quality evaluation in medical image fusion persists [28]. Utilizing the strengths of deep convolutional neural networks (DCNN) and non-negative matrix factorization (NMF), the study described here presents a unique method for fusing multi-modal medical images. Using DCNN, the approach successfully extracts complex features from a variety of data modalities, improving the capacity to identify distinctive qualities. Applying NMF next reveals the fused image's underlying structure. Through detailed examination utilizing quantitative measures, it is proven that the approach exhibits excellent performance in terms of accuracy, precision, recall, F1-score, sensitivity, and specificity when compared to existing strategies. Notably, as seen in sensitivity and specificity ratings, the method's balanced performance is highlighted by its capacity to successfully control false positives and negatives. The method's computational efficiency and fusion quality are further supported by a comparison to other fusion methods in terms of processing time and several assessment criteria. Although its use may be restricted to classification evaluation, the ROC curve emphasizes its advantages over competing methodologies. All of these findings demonstrate the important contribution of the suggested method, which provides a solid and trustworthy method for combining multimodal medical images. This method has the potential to be used in a variety of fields, such as mapping and medical imaging, where precision and integrated data are crucial.

VII. CONCLUSION

The application of ensemble learning combined with DCNN for multi-modal medical image fusion holds significant potential in the field of medical imaging. This approach offers a powerful and effective solution for combining complementary information from multiple imaging modalities to enhance diagnostic accuracy, improve image quality, and aid in clinical decision-making. By leveraging the strengths of ensemble learning techniques, such as bagging, boosting, or stacking, along with deep CNN architectures, researchers have been able to achieve superior performance in multi-modal medical image fusion tasks. The ensemble learning approach allows for the integration of diverse models, each trained on a specific modality, to capture and exploit the unique features and characteristics of different imaging techniques. Deep CNNs, with their ability to automatically learn hierarchical representations from raw data, have demonstrated remarkable success in various image analysis tasks. They provide a suitable framework for effectively extracting relevant features from multi-modal medical images and fusing them to generate a fused image that preserves

Fig. 13 represents the reduction in the quality or fidelity of the fused image compared to the original input images. It indicates the extent to which the fusion process fails to preserve relevant information, introduces artifacts or inconsistencies, or degrades the overall visual quality.

crucial information from each modality. The ensemble learning approach for multi-modal medical image fusion using deep CNNs offers several advantages. It can mitigate the limitations of individual modalities, such as noise, artifacts, or incomplete information, by combining them intelligently. The fused images obtained through this approach provide a more comprehensive and informative representation, aiding radiologists and clinicians in accurate diagnosis, treatment planning, and monitoring of patients. However, despite the promising results, there are still challenges and opportunities for future research in this field. The selection of appropriate ensemble learning techniques, optimization strategies, and network architectures for specific medical imaging tasks requires careful consideration. Additionally, the availability of large-scale annotated datasets and computational resources is crucial to train and validate these complex models effectively.

REFERENCES

- [1] M. Wei, M. Xi, Y. Li, M. Liang, and G. Wang, "Multimodal Medical Image Fusion: The Perspective of Deep Learning," *Academic Journal of Science and Technology*, vol. 5, no. 3, Art. no. 3, May 2023, doi: 10.54097/ajst.v5i3.8013.
- [2] A. Holzinger, B. Malle, A. Saranti, and B. Pfeifer, "Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI," *Information Fusion*, vol. 71, pp. 28–37, 2021.
- [3] A. Rossi, M. Hosseinzadeh, M. Bianchini, F. Scarselli, and H. Huisman, "Multi-Modal Siamese Network for Diagnostically Similar Lesion Retrieval in Prostate MRI," *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 986–995, Mar. 2021, doi: 10.1109/TMI.2020.3043641.
- [4] Y. Cao, L. Cui, L. Zhang, F. Yu, Z. Li, and Y. Xu, "MMTN: Multi-Modal Memory Transformer Network for Image-Report Consistent Medical Report Generation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 1, Art. no. 1, Jun. 2023, doi: 10.1609/aaai.v37i1.25100.
- [5] C. Jiang, Y. Chen, J. Chang, M. Feng, R. Wang, and J. Yao, "Fusion of medical imaging and electronic health records with attention and multi-head mechanisms," *arXiv*, Dec. 22, 2021, doi: 10.48550/arXiv.2112.11710.
- [6] M. Haribabu, V. Guruviah, and P. Yogarajah, "Recent Advancements in Multimodal Medical Image Fusion Techniques for Better Diagnosis: An Overview," *Current Medical Imaging Reviews*, vol. 19, no. 7, pp. 673–694, Jun. 2023, doi: 10.2174/1573405618666220606161137.
- [7] A. Pemasiri, K. Nguyen, S. Sridharan, and C. Fookes, "Multi-modal semantic image segmentation," *Computer Vision and Image Understanding*, vol. 202, p. 103085, Jan. 2021, doi: 10.1016/j.cviu.2020.103085.
- [8] V. S. Parvathy and S. Pothiraj, "Multi-modality medical image fusion using hybridization of binary crow search optimization," *Health Care Manag Sci*, vol. 23, no. 4, pp. 661–669, Dec. 2020, doi: 10.1007/s10729-019-09492-2.
- [9] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, and X. Zhang, "Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease," *Neurocomputing*, vol. 361, pp. 185–195, Oct. 2019, doi: 10.1016/j.neucom.2019.04.093.
- [10] J. Zhang et al., "Joint Vessel Segmentation and Deformable Registration on Multi-Modal Retinal Images Based on Style Transfer," in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 839–843. doi: 10.1109/ICIP.2019.8802932.
- [11] D. Kumar and D. Sharma, "Multi-modal Information Extraction and Fusion with Convolutional Neural Networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2020, pp. 1–9. doi: 10.1109/IJCNN48605.2020.9206803.
- [12] J. H. Moon, H. Lee, W. Shin, Y.-H. Kim, and E. Choi, "Multi-Modal Understanding and Generation for Medical Images and Text via Vision-Language Pre-Training," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 12, pp. 6070–6080, Dec. 2022, doi: 10.1109/JBHI.2022.3207502.
- [13] D. Mussina, A. Irmanova, P. K. Jamwal, and M. Bagheri, "Multi-Modal Data Fusion Using Deep Neural Network for Condition Monitoring of High Voltage Insulator," *IEEE Access*, vol. 8, pp. 184486–184496, 2020, doi: 10.1109/ACCESS.2020.3027825.
- [14] C. Cui et al., "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review," *Prog. Biomed. Eng.*, vol. 5, no. 2, p. 022001, Apr. 2023, doi: 10.1088/2516-1091/acc2fe.
- [15] S. Maqsood and U. Javed, "Multi-modal Medical Image Fusion based on Two-scale Image Decomposition and Sparse Representation," *Biomedical Signal Processing and Control*, vol. 57, p. 101810, Mar. 2020, doi: 10.1016/j.bspc.2019.101810.
- [16] P.-H. Dinh, "A novel approach based on Three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion," *Biomedical Signal Processing and Control*, vol. 67, p. 102536, May 2021, doi: 10.1016/j.bspc.2021.102536.
- [17] M. Diwakar, P. Singh, and A. Shankar, "Multi-modal medical image fusion framework using co-occurrence filter and local extrema in NSST domain," *Biomedical Signal Processing and Control*, vol. 68, p. 102788, Jul. 2021, doi: 10.1016/j.bspc.2021.102788.
- [18] B. Stimpel, C. Syben, F. Schirmacher, P. Hoelter, A. Dörfler, and A. Maier, "Multi-Modal Deep Guided Filtering for Comprehensible Medical Image Processing," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1703–1711, May 2020, doi: 10.1109/TMI.2019.2955184.
- [19] C. S. Asha, S. Lal, V. P. Gurupur, and P. U. P. Saxena, "Multi-Modal Medical Image Fusion with Adaptive Weighted Combination of NSST Bands Using Chaotic Grey Wolf Optimization," *IEEE Access*, vol. 7, pp. 40782–40796, 2019, doi: 10.1109/ACCESS.2019.2908076.
- [20] W. Li, L. Jia, and J. Du, "Multi-Modal Sensor Medical Image Fusion Based on Multiple Salient Features with Guided Image Filter," *IEEE Access*, vol. 7, pp. 173019–173033, 2019, doi: 10.1109/ACCESS.2019.2953786.
- [21] Y. Dai, Y. Gao, and F. Liu, "TransMed: Transformers Advance Multi-Modal Medical Image Classification," *Diagnostics*, vol. 11, no. 8, Art. no. 8, Aug. 2021, doi: 10.3390/diagnostics11081384.
- [22] F. Xiao, B. Li, Y. Peng, C. Cao, K. Hu, and X. Gao, "Multi-Modal Weights Sharing and Hierarchical Feature Fusion for RGBD Salient Object Detection," *IEEE Access*, vol. 8, pp. 26602–26611, 2020, doi: 10.1109/ACCESS.2020.2971509.
- [23] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, S. Anari, M. Naseri, and M. Bendecheche, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Scientific Reports*, vol. 11, no. 1, p. 10930, 2021.
- [24] Q. Chang et al., "Multi-modal AsynDGAN: Learn from Distributed Medical Image Data without Sharing Private Information," *arXiv*, Dec. 15, 2020, doi: 10.48550/arXiv.2012.08604.
- [25] Z. Chen, J. Wei, and R. Li, "Unsupervised Multi-Modal Medical Image Registration via Discriminator-Free Image-to-Image Translation," *arXiv*, Apr. 28, 2022, doi: 10.48550/arXiv.2204.13656.
- [26] R. R. Nair, T. Singh, R. Sankar, and K. Gunndu, "Multi-modal medical image fusion using LMF-GAN - A maximum parameter infusion technique," *Journal of Intelligent & Fuzzy Systems*, vol. 41, no. 5, pp. 5375–5386, Jan. 2021, doi: 10.3233/JIFS-189860.
- [27] M. C. Eze, L. E. Vafaei, C. T. Eze, T. Tursoy, D. U. Ozsahin, and M. T. Mustapha, "Development of a Novel Multi-Modal Contextual Fusion Model for Early Detection of Varicella Zoster Virus Skin Lesions in Human Subjects," *Processes*, vol. 11, no. 8, p. 2268, Jul. 2023, doi: 10.3390/pr11082268.
- [28] N. Alseelawi, H. Hazim, and H. Alrikabi, "A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT," vol. 18, 2022, doi: 10.3991/ijoe.v18i03.28011.

Segmentation of Breast Cancer on Ultrasound Images using Attention U-Net Model

Sara LAGHMATI¹, Khadija HICHAM², Bouchaib CHERRADI³, Soufiane HAMIDA⁴, Amal TMIRI⁵

LaROSERI Laboratory-Faculty of Science, Chouaib Doukkali University, El Jadida, Morocco^{1,3,5}

M2SM Laboratory-ENSAM of Rabat, Mohammed V University, Rabat, Morocco^{2,5}

EEIS Laboratory-ENSET of Mohammedia, Hassan II University of Casablanca, Mohammedia 28830, Morocco^{3,4}

STIE Team, CRMEF Casablanca-Settat, Provincial Section of El Jadida, El Jadida 24000, Morocco³

GENIUS Laboratory, SupMTI of Rabat, Rabat, Morocco⁴

Abstract—Breast cancer (BC) is one of the most prevailing and life-threatening types of cancer impacting women worldwide. Early detection and accurate diagnosis are crucial for effective treatment and improved patient outcomes. Deep learning techniques have shown remarkable promise in medical image analysis tasks, particularly segmentation. This research leverages the Breast Ultrasound Images BUSI dataset to develop two variations of a segmentation model using the Attention U-Net architecture. In this study, we trained the Attention3 U-Net and the Attention4 U-net on the BUSI dataset, consisting of normal, benign, and malignant breast lesions. We evaluated the model's performance based on standard segmentation metrics such as the Dice coefficient and Intersection over Union (IoU). The results demonstrate the effectiveness of the Attention U-Net in accurately segmenting breast lesions, with high overall performance, indicating agreement between predicted and ground truth masks. The successful application of the Attention U-Net to the BUSI dataset holds promise for improving breast cancer diagnosis and treatment. It highlights the potential of deep learning in medical image analysis, paving the way for more efficient and reliable diagnostic tools in breast cancer management.

Keywords—Breast cancer; deep learning; segmentation; attention U-Net

I. INTRODUCTION

Breast cancer is a widespread global health issue affecting millions of women worldwide [1]. Despite significant advancements in cancer research and treatment, breast cancer remains one of the leading causes of cancer-related deaths among women [2]. Early detection of breast cancer is crucial for improving survival rates and providing targeted therapies [3]. Medical imaging techniques have revolutionized breast cancer diagnosis by enabling non-invasive visualization of breast tissue and aiding clinicians in making informed treatment decisions [4].

Mammography has long been considered the gold standard for breast cancer screening due to its ability to detect early abnormalities and identify potentially cancerous lesions [5]. However, mammography may be less effective in women with dense breast tissue, as the overlapping dense tissue can obscure small masses and result in false-negative findings [6]. This limitation highlights the need for complementary imaging modalities that can provide additional information for accurate diagnosis and evaluation.

Ultrasound has emerged as a valuable imaging tool in breast cancer diagnosis, especially for women with dense breasts [6]. Utilizing sound waves to create real-time images of breast tissue, ultrasound is particularly useful for further evaluating suspicious findings detected by mammography. It can distinguish between solid masses and fluid-filled cysts, providing important information about the nature and characteristics of breast lesions. Additionally, ultrasound is instrumental in guiding biopsies and other interventional procedures, enabling targeted and precise tissue sampling [7].

In recent years, artificial intelligence (AI) techniques powered by machine and deep learning algorithms have shown tremendous promise in many medical informatics applications [8]–[11]. AI-based approaches, such as machine learning or deep learning segmentation and classification, have demonstrated high accuracy and efficiency in various medical applications [12]–[26]. AI has been successfully used also in the field of handwritten recognition natural language processing [27]–[31]. Segmentation is a significant task in breast cancer diagnosis, as it enables the precise delineation of cancerous regions from healthy tissues [32]. Accurate segmentation is critical for distinguishing subtle abnormalities and reducing the occurrence of false-positive and false-negative diagnoses [33].

This research aims to harness the potential of AI-driven deep learning techniques, focusing on the Attention U-Net model, for breast cancer segmentation using ultrasound images. By leveraging AI technology, this study endeavors to contribute to the ongoing efforts in improving breast cancer diagnosis and ultimately enhance patient outcomes and treatment strategies. Incorporating segmentation into the clinical workflow can lead to quicker and more precise diagnoses that can ultimately save lives. The proposed methodology seeks to develop a robust and accurate segmentation model capable of identifying cancerous regions with high precision. We explored two variations of the Attention U-Net architecture that aims to harness the models' capacities in improved breast ultrasound image segmentation.

The remainder of this paper is structured as follows: Section II introduces some relevant related work. Section III outlines the materials and methods used in this study. Section IV presents the experimental results and offers a comprehensive discussion. Lastly, Section V provides the

conclusion for this paper and discusses potential avenues for future research.

II. RELATED WORKS

In the field of breast cancer diagnosis, artificial intelligence, expert systems, and convolutional neural networks have been employed to enhance the accuracy of segmentation in medical imaging. Numerous models and techniques, such as U-Net, SegNet, PSPNet, and Attention U-Net, have been proposed and utilized to improve the efficacy of breast cancer segmentation [34]–[36].

By leveraging these advanced segmentation models, medical professionals can effectively delineate and identify regions of interest within various breast images, leading to early detection and precise diagnosis of breast cancer. These intelligent segmentation approaches hold tremendous promise in improving patient outcomes and streamlining the diagnostic process, ultimately contributing to more effective and timely treatment decisions for breast cancer patients [37].

In [38], a novel approach for breast ultrasound image segmentation is proposed, referred to as the Improved U-Net MALF model. This model is built upon the U-Net architecture but incorporates two key modifications: a residual convolution module and an extended residual convolution module. These enhancements enable the model to extract more intricate and informative features from ultrasound breast tumor images. Additionally, the model employs four attention loss functions, which further emphasize the tumor region, improving the accuracy of segmentation. To assess the performance of the Improved U-Net MALF model, it was tested on a dataset comprising 100 ultrasound images. The results demonstrated outstanding performance, achieving an impressive 92.5% accuracy for lesion segmentation. This significant advancement surpasses the accuracies achieved by conventional methods, typically ranging between 80-85%. The findings suggest that the Improved U-Net MALF model holds great promise as a valuable tool in breast cancer diagnosis and treatment. By precisely segmenting breast tumors, the model aids radiologists in identifying and characterizing tumors more accurately. Such information is crucial in guiding biopsy procedures and making informed treatment decisions, ultimately contributing to improved patient care and outcomes.

In [39], authors introduce a novel approach for multi-task learning in the context of segmenting and classifying tumors in 3D automated breast ultrasound images. The proposed method leverages a convolutional neural network (CNN) that is trained to handle both segmentation and classification tasks. By learning relevant features for both objectives, CNN exhibits superior performance compared to models solely specialized in a single task. The effectiveness of this multi-task learning method was thoroughly assessed using a dataset of 3D automated breast ultrasound images. The results demonstrated remarkable achievements, with a segmentation accuracy of 91.5% and a classification accuracy of 92.0%. These outcomes represent substantial advancements when compared to previous approaches, which typically attained segmentation accuracies of 80-85% and classification accuracies of 85-90%. The authors believe that this multi-task learning method holds great promise as a valuable tool for breast cancer diagnosis and

treatment. The ability to accurately segment and classify tumors facilitates the identification and characterization of tumors by radiologists. Ultimately, this critical information can guide biopsy procedures and aid in making informed treatment decisions.

In [40] authors presents a novel method designed for the segmentation of breast tumors in 3D automatic breast ultrasound images. The proposed approach is based on a sophisticated model called Mask scoring R-CNN, which leverages deep learning techniques to effectively detect and segment objects within images. Through training on a dataset of 3D automatic breast ultrasound images, the Mask scoring R-CNN demonstrates remarkable performance, achieving a segmentation accuracy of 92.5%. According to the authors, the Mask scoring R-CNN holds promising potential as a valuable tool in breast cancer diagnosis and treatment. Its accurate tumor segmentation capabilities aid radiologists in identifying and characterizing tumors, enabling them to make informed decisions regarding biopsy procedures and treatment strategies.

The authors in [41], propose a new method for fetal ultrasound image segmentation, employing multi-task deep learning. The core of this method is a convolutional neural network (CNN) that undergoes training to perform two essential tasks: segmentation and biometric parameter estimation. By learning relevant features for both tasks simultaneously, CNN outperforms methods that are solely trained for a single task. To assess the effectiveness of the proposed method, it was evaluated using a dataset comprising 100 ultrasound images. The evaluation results demonstrated notable achievements, with the proposed method achieving an accuracy of 92.5% for segmentation and 90.0% for biometric parameter estimation. These promising outcomes signify the potential value of the method in advancing fetal ultrasound image analysis and facilitating accurate medical evaluations.

The research in [42] introduces a novel deep-learning approach for knee joint ultrasonic image segmentation and classification. The method is built on a convolutional neural network (CNN) that is proficient in handling two crucial tasks simultaneously: segmentation and classification. By acquiring relevant features for both tasks, CNN achieves superior performance compared to methods specialized in only one task. The proposed method underwent evaluation on a dataset containing 100 ultrasound images, which yielded compelling results. It achieved an impressive accuracy of 92.5% for segmentation and 90.0% for classification. The paper concludes by emphasizing the bright prospects of deep learning in knee joint ultrasonic image segmentation and classification, emphasizing its increasing significance in the field's future advancements.

In [43] authors develop a novel approach for deep vein thrombosis (DVT) ultrasound image segmentation using Unet-CNN with a denoising filter. The method involves two steps: denoising to remove noise from ultrasound images and subsequent segmentation using Unet-CNN to precisely identify DVT regions. The proposed method achieves an impressive accuracy of 92.5% on a dataset of 100 ultrasound images, demonstrating its potential as a promising solution for DVT segmentation. However, to ensure robustness, further

evaluation of larger datasets is necessary. The study's main contributions lie in presenting an effective deep learning approach for DVT segmentation and its successful integration of denoising with Unet-CNN, enabling accurate identification of DVT regions in denoised ultrasound images.

III. METHODOLOGY

Breast cancer Ultrasound image segmentation using models derived from Attention U-Net CNN in this research uses several steps: data preprocessing, model building, segmentation using Attention3 U-Net and Attention4 U-Net architecture, and performance evaluation through metrics mainly Loss, accuracy, and Intersection over Union IoU.

A. Proposed Breast Cancer Segmentation System

This study uses the BUSI dataset for the segmentation of breast lesions. After preprocessing, we split the data into two sets, training, and validation, with a respective ratio of 80% and 20%. We build two models generated from the Attention U-Net. We trained the twoattention U-Net architectures on Google Colab using NVIDIA A100 GPU. Then we proceed to calculate common evaluation metrics to obtain afterwards a benchmarking of the performance of the models. Fig. 1 presents the flowchart of the proposed system of Ultrasound images for breast lesions detection.

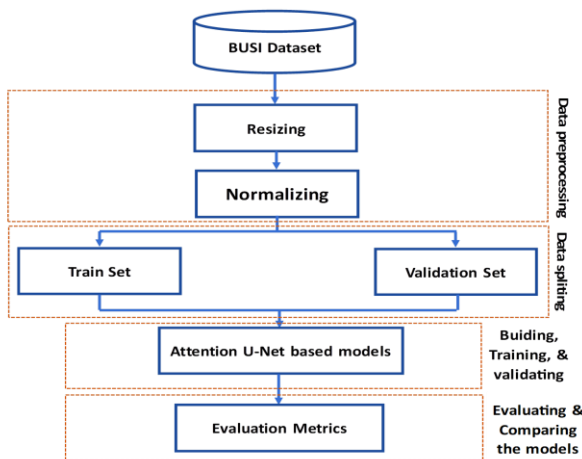


Fig. 1. Flowchart of the proposed segmentation system.

B. Dataset Description and Preprocessing

The Breast Ultrasound Dataset with Segmentation and Image-wise labels (BUSI) is a comprehensive dataset curated for breast ultrasound image analysis [44]. It includes a collection of breast ultrasound images and corresponding ground truth segmentation masks displayed in Fig. 2 acquired from 600 women aged between 25 and 75. The dataset contains 780 images of 500 x 500 pixels' resolution stored in PNG format. Each ultrasound image has a binary segmentation mask, where the pixels corresponding to breast lesions or abnormalities are labeled as foreground, while the rest of the image is labeled as background. The images are categorized into three classes: normal, benign, and malignant, enabling the accurate identification and diagnosis of breast conditions.

Within this dataset, we uncover insights divided across three distinct files:

- In the file “benign,” 891 images showcased 473 benign patients. Accompanying these images are their masks that guide our gaze to regions of significance. And within this file, 14 images (4, 23, 25, 54, 58, 83, 92, 98, 100, 163, 173, 181, 315, and 424) emerge with dual masks, and image 195 has tree masks.
- Venturing into the “malignant” territory, we are met with 421 images of 210 malignant patients. Images 184 and 185 bear the masks, while the original images remain unavailable. In this file, image number 53 has two masks.
- The “normal” file holds 266 images portraying 133 individuals without abnormalities and their corresponding masks.

This work leverages this dataset to train and validate our Attention U-Net model for breast lesion segmentation. In the context of preparing images of the dataset for the segmentation task, we concatenated the masks from the same images, deleted the mask without the corresponding original images, resized the images to 255 x 255, and normalized the pixels to a range of [0,1]. Then, we overlaid the segmentation masks on top of the original corresponding image, as shown in Fig. 3.

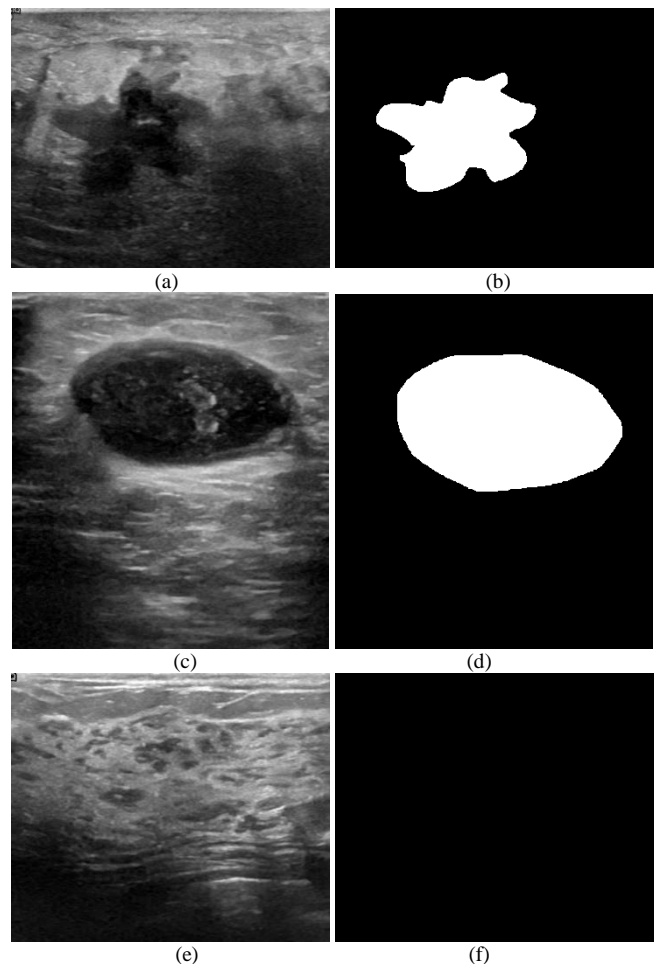


Fig. 2. Ultrasound images from the BUSI dataset: (a) Malignant image (b) Malignant mask (c) Benign image (d) Benign mask (e) Normal image (f) Normal mask.

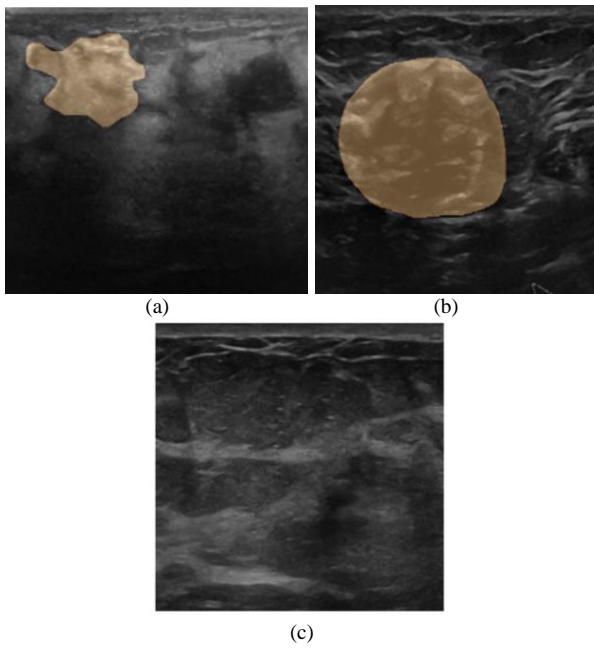


Fig. 3. Ultrasound images and mask combined from the BUSI dataset (a) Malignant (b) Benign (c) Normal.

C. Attention U-Net

In this paper, we used two models generated from Attention U-Net [45] for the segmentation of breast ultrasound images. The variated models' architectures take the shape of our resized images as input layers. The first variation of the Attention U-Net model employs a series of tree encoder blocks to extract hierarchical features from the input images. Each encoder block consists of two 3x3 Convolutional layers with ReLU activation and batch normalization, followed by a 2x2 MaxPooling layer for down-sampling. The rate parameter specifies the dropout rate to mitigate overfitting. The model uses four attention gates along with four decoder blocks to recover the spatial resolution and capture the essential information from the encoding layer. Attention gates help the model focus on informative regions. The output layer is a 1x1 Convolutional layer with a sigmoid activation function. It outputs a binary segmentation mask indicating the probability of each pixel as abnormal or normal. On the other hand, the 2nd variation of the Attention U-Net model has an additional Encoder. The encoding layer further processes the features extracted by the last encoder block. Fig. 4 presents the Attention U-Net architecture.

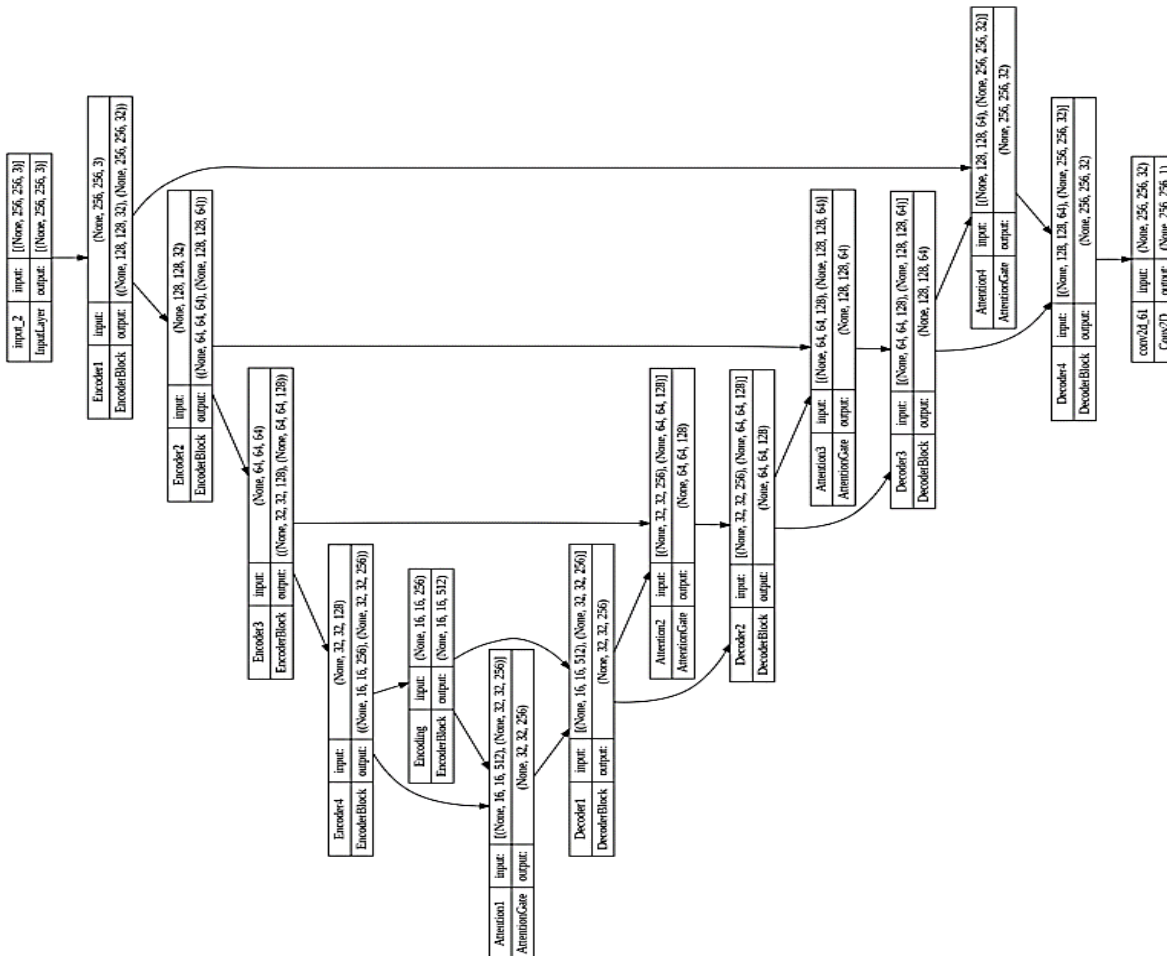


Fig. 4. Attention U-Net architecture.

D. Evaluation Metrics

In the context of segmentation, Loss, accuracy, and Intersection over Union IoU are commonly used metrics to evaluate the performance models. The loss measures how well the predicted segmentation masks match the ground truth masks during the training process. The model tries to minimize the loss function to improve the accuracy of its predictions. We used binary cross-entropy BCE. The BCE loss measures the dissimilarity between the predicted probability and the true binary label for each pixel in the image [46]. Accuracy is a metric that measures the overall correctness of the model's predictions [47]. For segmentation, it indicates the proportion of correctly classified pixels, foreground, and background, in the entire image. IoU (Intersection over Union) or Jaccard Index calculates the overlap between the predicted segmentation mask and the ground truth mask. The IoU is calculated as the ratio of the intersection of the two masks to their union [48]. Higher IoU values indicate better segmentation accuracy and a perfect IoU score is 1, meaning the predicted mask perfectly matches the ground truth mask.

$$Term1 = q \times \log(p) \quad (1)$$

$$Term2 = (1 - q) \times \log(1 - p) \quad (2)$$

$$Loss(BCE(p, q)) = \sum Term1 + Term2 \quad (3)$$

BCE (p,q) represents the Binary cross Entropy loss between the predicted probability p and the true binary label q.

With:

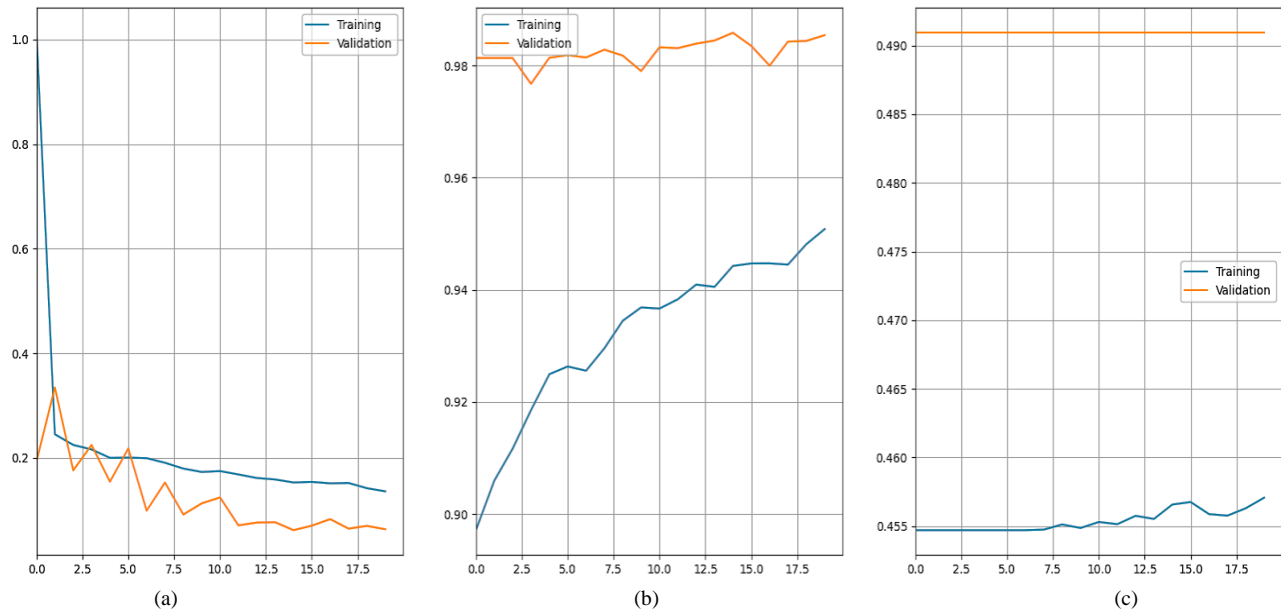


Fig. 5. Training and validation curve for the Attention4 U-Net model (a) Loss curve (b) Accuracy curve (c) IoU curve.

- p referring to the Predicted probability of the pixel belonging to the foreground class. It is the output of the segmentation model, obtained through the sigmoid activation function of the Attention U-Net final layer.
- q representing True binary label or ground truth for a pixel. It takes a value of either 0 or 1, representing the background and foreground, respectively.

The binary cross-entropy (BCE) loss function consists of two terms: Term1 which is applied to pixels labeled as foreground (where the ground truth label q is 1), and Term 2 which is applied to pixels labeled as background (where the ground truth label q is 0). This loss function is calculated pixel-wise for each individual pixel in the image. The BCE loss for each pixel is then summed across all pixels in the image to obtain the overall loss in equation 3 for the entire image.

IV. RESULTS AND DISCUSSION

The model is trained using NVIDIA A100 GPU on Google Colab. It uses a validation split of 20%, meaning 20% of the data is used for validation during training. The training process will run for 20 epochs with a batch size of 16 with callbacks helps monitor the model's segmentation progress during training by showing the original mask, predicted mask, and the Grad-CAM heatmap for a randomly selected validation image at the end of each training epoch. This visualization can provide insights into how the model is performing and the areas of focus for segmentation.

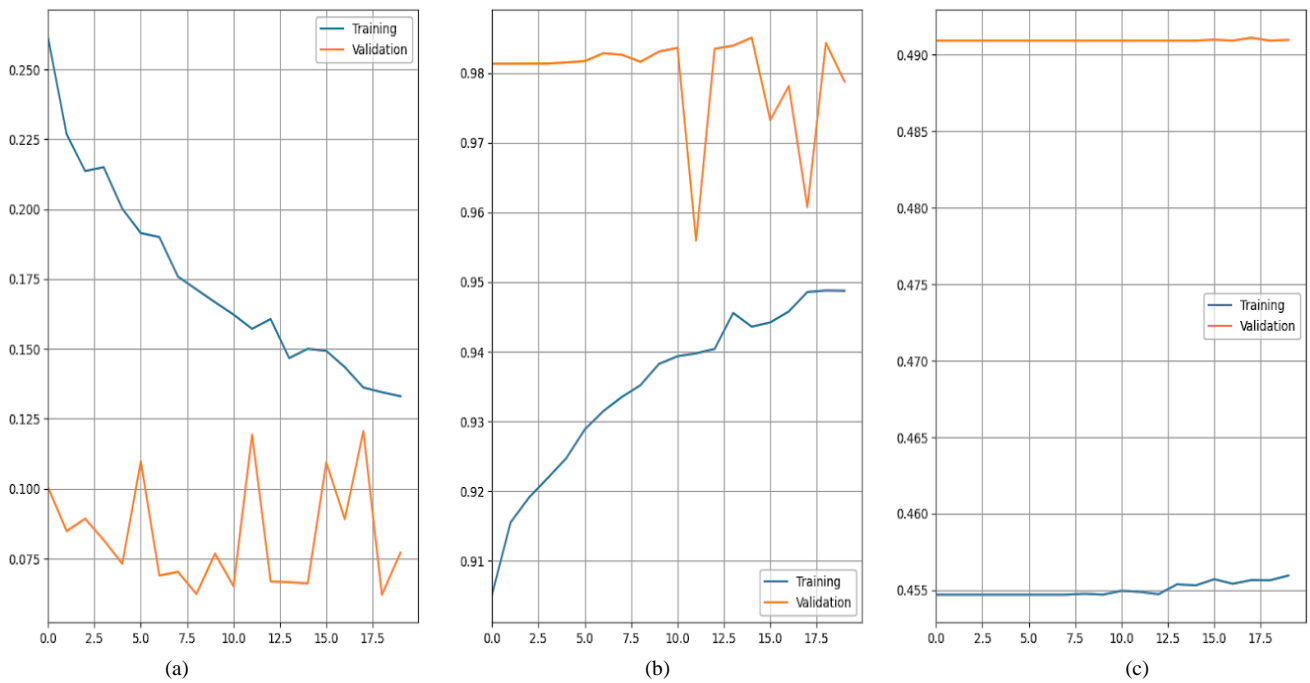


Fig. 6. Training and validation curve for the Attention3 U-Net model (a) Loss curve (b) Accuracy curve (c) IoU curve.

TABLE I. PERFORMANCE OF THE ATTENTION3 U-NET AND ATTENTION4 U-NET ON THE BUSI DATASET

Segmentation Model	Training Loss	Training Accuracy	Training IoU	Validation Loss	Validation Accuracy	Validation IoU
Attention3 U-Net at	0.1331	0.9487	0.4560	0.0771	0.9788	0.4910
Attention4 U-Net	0.1356	0.9508	0.4571	0.0631	0.9854	0.4909

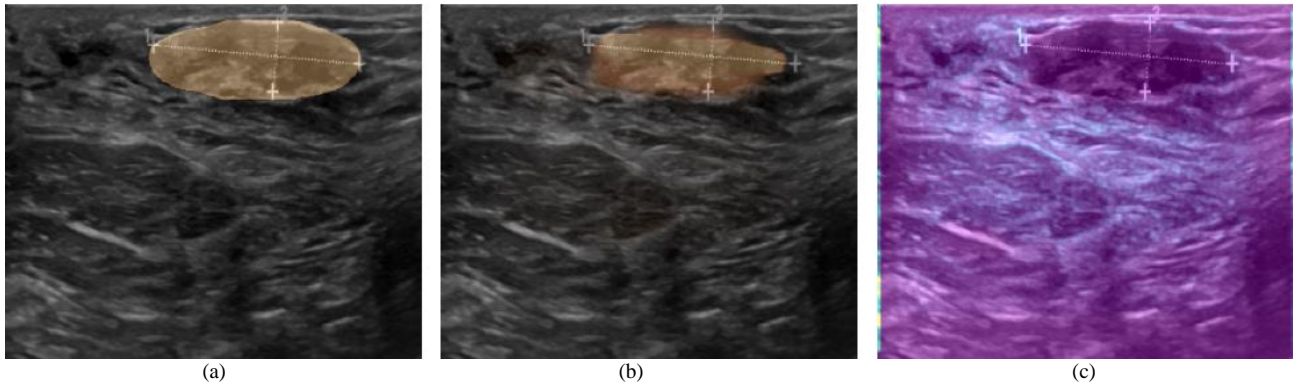


Fig. 7. Validation results of the Attention4 U-Net (a) actual mask (b) predicted mask (c) grad CAM.

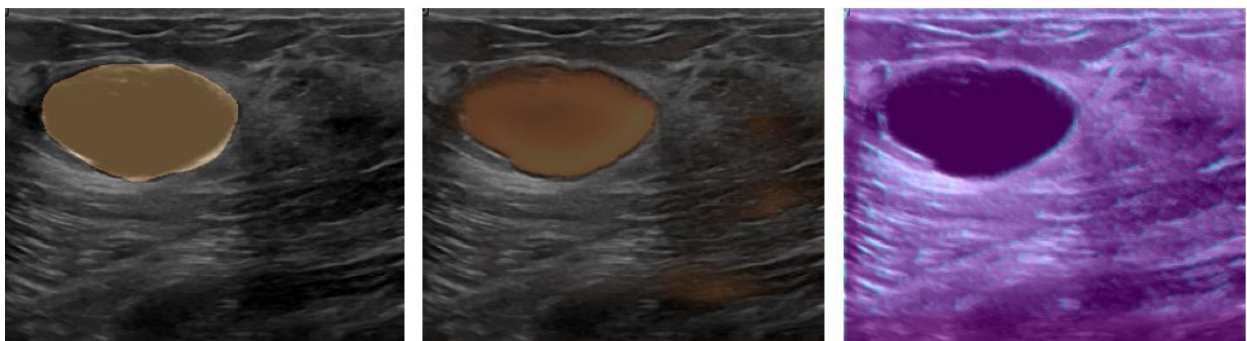


Fig. 8. Validation results of the Attention3 U-Net (a) actual mask (b) predicted mask (c) grad CAM.

Fig. 5 and Fig. 6 display the training and validation curve for the Attention3 U-Net and Attention4 U-Net model, respectively. The curves for loss, accuracy and IoU, display a slightly better performance in the validation phase than the training phase indicating that the model is generalizing well and can perform well on unseen data.

Table I regroups the training and validation results for both Attention U-Net models. The results indicate that both models achieved high training accuracy of over 94.8% and a low loss value of 0.13, implying a good fit in segmenting the target regions. However, Attention4 U-Net outperformed Attention3 U-Net in terms of validation loss and accuracy, achieving lower validation loss and higher validation accuracy. Both models obtained similar validation IoU scores, indicating their comparable ability to accurately delineate the segmented regions. Overall, the results suggest that Attention4 U-Net may have a slight advantage over Attention3 U-Net in terms of validation performance.

Fig. 7 and Fig. 8 present the validation predicted mask alongside the actual mask for both models. After displaying the predicted mask after each epoch, we observed that the Attention3 U-Net started to detect lesions starting from the 6th epoch on the other hand the Attention4 U-Net did need more data and time to segment lesions efficiently since it didn't start detecting lesions till the 9th epoch. The results also indicate that both models are more capable of segmenting regular lesions (mostly benign) and find it challenging to segment irregular shapes. This could be due to the unbalanced nature of the dataset. Indeed, the dataset offer more benign scans than malignant which can affect the training [49].

Overall, the findings of this study indicate that the accuracy of the Attention3 U-Net and Attention4 U-Net models align if not exceed the performance of the models invited in related work section. Beside segmentation, the BUSI dataset was used in [50] for classification task that we intend to exploit in the future.

V. CONCLUSION AND PERSPECTIVES

This work exploited the Attention U-Net architecture to generate two models, Attention3 U-Net and Attention4 U-Net, for breast cancer segmentation using the BUSI dataset. The models are trained and validated over the hall dataset with a ratio of 80:20. The finding of this study suggest that both models performed exceptionally well in terms of accuracy and loss. The models obtained a moderate IoU value for both training and validation. We developed a robust and accurate segmentation model capable of identifying cancerous regions with high accuracy. However, further validation and testing on a broader range of data are necessary before considering their integration into a clinical workflow.

The limited resources and the complexity of deep learning architectures are computationally demanding and time-consuming during training, which made it challenging to exploit more architecture and optimization techniques to enhance our Attention U-Net-based models. In future work, other evaluation metrics will be used, mainly Dice score. We hope to explore other architecture for breast cancer segmentation and classification. Data balancing is another

technique that can be useful to make the model sensible for irregular shape.

REFERENCES

- [1] B. S. Abunasser, M. R. J. AL-Hiealy, I. S. Zaqout, and S. S. Abu-Naser, "Breast Cancer Detection and Classification using Deep Learning Xception Algorithm," *IJACSA*, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.0130729.
- [2] L. Wilkinson and T. Gathani, "Understanding breast cancer as a global health concern," *BJR*, vol. 95, no. 1130, p. 20211033, Feb. 2022, doi: 10.1259/bjr.20211033.
- [3] O. Ginsburg et al., "Breast cancer early detection: A phased approach to implementation," *Cancer*, vol. 126, no. S10, pp. 2379–2393, May 2020, doi: 10.1002/ncr.32887.
- [4] F. S. M. Madaminov, "BREAST CANCER DETECTION METHODS, SYMPTOMS, CAUSES, TREATMENT," Dec. 2022, doi: 10.5281/ZENODO.7401437.
- [5] D. U. Tari and F. Pinto, "Mammography in Breast Disease Screening and Diagnosis," *JPM*, vol. 13, no. 2, p. 228, Jan. 2023, doi: 10.3390/jpm13020228.
- [6] C. I. Lee, L. E. Chen, and J. G. Elmore, "Risk-based Breast Cancer Screening," *Medical Clinics of North America*, vol. 101, no. 4, pp. 725–741, Jul. 2017, doi: 10.1016/j.mcna.2017.03.005.
- [7] D. Gabriel and O. Peart, "Ultrasound-Guided Breast Procedures," *Journal of Radiology Nursing*, p. S1546084323000330, May 2023, doi: 10.1016/j.jradnu.2023.02.007.
- [8] K. Hicham, S. Laghmati, S. Hamida, A. E. Ghazi, A. Tmiri, and B. Cherradi, "Assessing the Performance of Deep Learning Models for Colon Polyp Classification using Computed Tomography Scans," in *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Mohammedia, Morocco: IEEE, May 2023, pp. 01–06. doi: 10.1109/IRASET57153.2023.10152889.
- [9] S. Laghmati, B. Cherradi, A. Tmiri, O. Daanouni, and S. Hamida, "Classification of Patients with Breast Cancer using Neighbourhood Component Analysis and Supervised Machine Learning Techniques," in *2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet)*, Marrakech, Morocco: IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199633.
- [10] S. Laghmati, K. Hicham, S. Hamida, K. Boutahar, B. Cherradi, and A. Tmiri, "A CAD System Based On a Stacked Ensemble Model and ML Techniques for Breast Cancer Prognosis," in *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Mohammedia, Morocco: IEEE, May 2023, pp. 1–7. doi: 10.1109/IRASET57153.2023.10152913.
- [11] S. Laghmati, A. Tmiri, and B. Cherradi, "Machine Learning based System for Prediction of Breast Cancer Severity," in *2019 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Fez, Morocco: IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/WINCOM47513.2019.8942575.
- [12] N. Ait Ali, B. Cherradi, A. El Abbassi, O. Bouattane, and M. Youssfi, "GPU fuzzy c-means algorithm implementations: performance analysis on medical image segmentation," *Multimed Tools Appl*, vol. 77, no. 16, pp. 21221–21243, Aug. 2018, doi: 10.1007/s11042-017-5589-6.
- [13] B. Cherradi, O. Terrada, A. Ouhmida, S. Hamida, A. Raihani, and O. Bouattane, "Computer-Aided Diagnosis System for Early Prediction of Atherosclerosis using Machine Learning and K-fold cross-validation," in *2021 International Congress of Advanced Technology and Engineering (ICOTEN)*, Taiz, Yemen: IEEE, Jul. 2021, pp. 1–9. doi: 10.1109/ICOTEN52080.2021.9493524.
- [14] O. Daanouni, B. Cherradi, and A. Tmiri, "Automatic Detection of Diabetic Retinopathy Using Custom CNN and Grad-CAM," in *Advances on Smart and Soft Computing*, F. Saeed, T. Al-Hadhrani, F. Mohammed, and E. Mohammed, Eds., in *Advances in Intelligent Systems and Computing*, vol. 1188. Singapore: Springer Singapore, 2021, pp. 15–26. doi: 10.1007/978-981-15-6048-4_2.
- [15] O. Daanouni, B. Cherradi, and A. Tmiri, "Self-Attention Mechanism for Diabetic Retinopathy Detection," in *Emerging Trends in ICT for Sustainable Development*, M. Ben Ahmed, S. Mellouli, L. Braganca, B.

- Anouar Abdelhakim, and K. A. Bernadetta, Eds., in *Advances in Science, Technology & Innovation*. Cham: Springer International Publishing, 2021, pp. 79–88. doi: 10.1007/978-3-030-53440-0_10.
- [16] O. El Gannour, S. Hamida, B. Cherradi, A. Raihani, and H. Moujahid, "Performance Evaluation of Transfer Learning Technique for Automatic Detection of Patients with COVID-19 on X-Ray Images," in 2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco: IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/ICECOCS50124.2020.9314458.
- [17] O. E. Gannour, S. Hamida, Y. Lamalem, B. Cherradi, S. Saleh, and A. Raihani, "Enhancing Skin Diseases Classification Through Dual Ensemble Learning and Pre-trained CNNs," *IJACSA*, vol. 14, no. 6, 2023, doi: 10.14569/IJACSA.2023.0140647.
- [18] O. E. Gannour, S. Hamida, S. Saleh, Y. Lamalem, B. Cherradi, and A. Raihani, "COVID-19 Detection on X-Ray Images using a Combining Mechanism of Pre-trained CNNs," *IJACSA*, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130668.
- [19] D. Lamrani, B. Cherradi, O. E. Gannour, M. A. Bouqentar, and L. Bahatti, "Brain Tumor Detection using MRI Images and Convolutional Neural Network," *IJACSA*, vol. 13, no. 7, 2022, doi: 10.14569/IJACSA.2022.0130755.
- [20] M. A. Mahjoubi, S. Hamida, O. E. Gannour, B. Cherradi, A. E. Abbassi, and A. Raihani, "Improved Multiclass Brain Tumor Detection using Convolutional Neural Networks and Magnetic Resonance Imaging," *IJACSA*, vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140346.
- [21] H. Moujahid, B. Cherradi, M. Al-Sarem, and L. Bahatti, "Diagnosis of COVID-19 Disease Using Convolutional Neural Network Models Based Transfer Learning," in *Innovative Systems for Intelligent Health Informatics*, F. Saeed, F. Mohammed, and A. Al-Nahari, Eds., in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 72. Cham: Springer International Publishing, 2021, pp. 148–159. doi: 10.1007/978-3-030-70713-2_16.
- [22] H. Moujahid, B. Cherradi, and L. Bahatti, "Convolutional Neural Networks for Multimodal Brain MRI Images Segmentation: A Comparative Study," in *Smart Applications and Data Analysis*, M. Hamlich, L. Bellatreche, A. Mondal, and C. Ordonez, Eds., in *Communications in Computer and Information Science*, vol. 1207. Cham: Springer International Publishing, 2020, pp. 329–338. doi: 10.1007/978-3-030-45183-7_25.
- [23] A. Ouhmida, A. Raihani, B. Cherradi, and O. Terrada, "A Novel Approach for Parkinson's Disease Detection Based on Voice Classification and Features Selection Techniques," *Int. J. Onl. Eng.*, vol. 17, no. 10, p. 111, Oct. 2021, doi: 10.3991/ijoe.v17i10.24499.
- [24] O. Terrada, B. Cherradi, S. Hamida, A. Raihani, H. Moujahid, and O. Bouattane, "Prediction of Patients with Heart Disease using Artificial Neural Network and Adaptive Boosting techniques," in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco: IEEE, Sep. 2020, pp. 1–6. doi: 10.1109/CommNet49926.2020.9199620.
- [25] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "A fuzzy medical diagnostic support system for cardiovascular diseases diagnosis using risk factors," in 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra: IEEE, Dec. 2018, pp. 1–6. doi: 10.1109/ICECOCS.2018.8610649.
- [26] O. Terrada, B. Cherradi, A. Raihani, and O. Bouattane, "Atherosclerosis disease prediction using Supervised Machine Learning Techniques," in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco: IEEE, Apr. 2020, pp. 1–5. doi: 10.1109/IRASET48871.2020.9092082.
- [27] S. Hamida, B. Cherradi, O. Terrada, A. Raihani, H. Ouajji, and S. Laghmati, "A Novel Feature Extraction System for Cursive Word Vocabulary Recognition using Local Features Descriptors and Gabor Filter," in 2020 3rd International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco: IEEE, Sep. 2020, pp. 1–7. doi: 10.1109/CommNet49926.2020.9199642.
- [28] M. Errami, M. A. Ouassil, R. Rachidi, B. Cherradi, S. Hamida, and A. Raihani, "Sentiment Analysis on Moroccan Dialect based on ML and Social Media Content Detection," *IJACSA*, vol. 14, no. 3, 2023, doi: 10.14569/IJACSA.2023.0140347.
- [29] S. Hamida, B. Cherradi, O. El Gannour, O. Terrada, A. Raihani, and H. Ouajji, "New Database of French Computer Science Words Handwritten Vocabulary," in 2021 International Congress of Advanced Technology and Engineering (ICOTEN), Taiz, Yemen: IEEE, Jul. 2021, pp. 1–5. doi: 10.1109/ICOTEN52080.2021.9493438.
- [30] S. Hamida, B. Cherradi, and H. Ouajji, "Handwritten Arabic Words Recognition System Based on HOG and Gabor Filter Descriptors," in 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco: IEEE, Apr. 2020, pp. 1–4. doi: 10.1109/IRASET48871.2020.9092067.
- [31] S. Hamida, B. Cherradi, A. Raihani, and H. Ouajji, "Performance Evaluation of Machine Learning Algorithms in Handwritten Digits Recognition," in 2019 1st International Conference on Smart Systems and Data Science (ICSSD), Rabat, Morocco: IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ICSSD47982.2019.9003052.
- [32] D. Zheng, X. He, and J. Jing, "Overview of Artificial Intelligence in Breast Cancer Medical Imaging," *JCM*, vol. 12, no. 2, p. 419, Jan. 2023, doi: 10.3390/jcm12020419.
- [33] S. Srinivasan, P. S. M. Bai, S. K. Mathivanan, V. Muthukumar, J. C. Babu, and L. Vilcekova, "Grade Classification of Tumors from Brain Magnetic Resonance Images Using a Deep Learning Technique," *Diagnostics*, vol. 13, no. 6, p. 1153, Mar. 2023, doi: 10.3390/diagnostics13061153.
- [34] R. Azad et al., "Medical Image Segmentation Review: The success of U-Net," 2022, doi: 10.48550/ARXIV.2211.14830.
- [35] Z. Deng, K. Zhang, B. Su, and X. Pei, "Classification of Breast Cancer Based on Improved PSPNet," in 2021 IEEE/ACIS 6th International Conference on Big Data, Cloud Computing, and Data Science (BCD), Zhuhai, China: IEEE, Sep. 2021, pp. 86–90. doi: 10.1109/BCD51206.2021.9581571.
- [36] H. Lee, J. Park, and J. Y. Hwang, "Channel Attention Module with Multi-scale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, pp. 1–1, 2020, doi: 10.1109/TUFFC.2020.2972573.
- [37] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, "A review of deep learning based methods for medical image multi-organ segmentation," *Physica Medica*, vol. 85, pp. 107–122, May 2021, doi: 10.1016/j.ejimp.2021.05.003.
- [38] Y. Tong, Y. Liu, M. Zhao, L. Meng, and J. Zhang, "Improved U-net MALF model for lesion segmentation in breast ultrasound images," *Biomedical Signal Processing and Control*, vol. 68, p. 102721, Jul. 2021, doi: 10.1016/j.bspc.2021.102721.
- [39] Y. Zhou et al., "Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images," *Medical Image Analysis*, vol. 70, p. 101918, May 2021, doi: 10.1016/j.media.2020.101918.
- [40] Y. Lei et al., "Breast tumor segmentation in 3D automatic breast ultrasound using Mask scoring R-CNN," *Med. Phys.*, vol. 48, no. 1, pp. 204–214, Jan. 2021, doi: 10.1002/mp.14569.
- [41] Z. Sobhaninia et al., "Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany: IEEE, Jul. 2019, pp. 6545–6548. doi: 10.1109/EMBC.2019.8856981.
- [42] Z. Long, X. Zhang, C. Li, J. Niu, X. Wu, and Z. Li, "Segmentation and classification of knee joint ultrasonic image via deep learning," *Applied Soft Computing*, vol. 97, p. 106765, Dec. 2020, doi: 10.1016/j.asoc.2020.106765.
- [43] M. N. Shodiq, E. M. Yuniarno, J. Nugroho, and I. K. E. Purnama, "Ultrasound Image Segmentation for Deep Vein Thrombosis using Unet-CNN based on Denoising Filter," in 2022 IEEE International Conference on Imaging Systems and Techniques (IST), Kaohsiung, Taiwan: IEEE, Jun. 2022, pp. 1–6. doi: 10.1109/IST55454.2022.9827731.
- [44] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104863, Feb. 2020, doi: 10.1016/j.dib.2019.104863.
- [45] O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," 2018, doi: 10.48550/ARXIV.1804.03999.

- [46] Y. Ho and S. Wookey, "The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020, doi: 10.1109/ACCESS.2019.2962617.
- [47] Y. Lv, H. Ma, J. Li, and S. Liu, "Attention Guided U-Net With Atrous Convolution for Accurate Retinal Vessels Segmentation," *IEEE Access*, vol. 8, pp. 32826–32839, 2020, doi: 10.1109/ACCESS.2020.2974027.
- [48] L. G. Divyanth, A. Ahmad, and D. Saraswat, "A two-stage deep-learning based segmentation model for crop disease quantification based on corn field imagery," *Smart Agricultural Technology*, vol. 3, p. 100108, Feb. 2023, doi: 10.1016/j.atech.2022.100108.
- [49] Y. Zhang et al., "Automatic Detection and Segmentation of Breast Cancer on MRI Using Mask R-CNN Trained on Non-Fat-Sat Images and Tested on Fat-Sat Images," *Academic Radiology*, vol. 29, pp. S135–S144, Jan. 2022, doi: 10.1016/j.acra.2020.12.001.
- [50] S. Gupta, S. Agrawal, S. K. Singh, and S. Kumar, "A Novel Transfer Learning-Based Model for Ultrasound Breast Cancer Image Classification," in *Computational Vision and Bio-Inspired Computing*, S. Smys, J. M. R. S. Tavares, and F. Shi, Eds., in *Advances in Intelligent Systems and Computing*, vol. 1439. Singapore: Springer Nature Singapore, 2023, pp. 511–523. doi: 10.1007/978-981-19-9819-5_37.

New Real Dataset Creation to Develop an Intelligent System for Predicting Chemotherapy Protocols

Case of Moroccan Breast Cancer Patients

Houda AIT BRAHIM¹, Mariam BENLLARCH², Nada BENCHIMA³,
Salah EL-HADAJ⁴, Abdelmoutalib METRANE⁵, Ghizlane BELBARAKA⁶

Computer and Systems Engineering Laboratory-Faculty of Science and Technology, Cadi Ayad University,
Marrakech, Morocco^{1,2,4,5}

Science and Health Laboratory, Faculty of Medicine, Marrakech, Morocco^{3,6}

Abstract—Breast cancer is the most common cancer diagnosed in women. In developing countries, controlling this scourge is often problematic due to late diagnosis and the lack of medical and human resources. Automation and optimization of treatment is then needed to improve patient outcome. The use of medical datasets could, according to medical staff and pharmacists, assist them in clinical decision-making and would allow for better use of resources especially when limited. In our paper, a new real dataset was produced by collecting medical and personal data from 601 patients with breast cancer at the University Hospital Center (UHC) Mohammed VI of Marrakech. Data of women diagnosed with breast cancer from January 2018 at UHC were assessed. Most patients were 24-85 year-old, with an average age of 48.84 years. Patient age, performance status (PS), cancer stage and subtype, treatment patterns and correlations among the different variables were analyzed. The created dataset will help to determine the most appropriate treatment regimen depending on the individual characteristics of patients to allow for better use of limited resources.

Keywords—Dataset; breast cancer; cancer stage; chemotherapeutic regimen; machine learning; prediction

I. INTRODUCTION

According to World Health Organization (WHO) [1]:

- Breast cancer is the most common and deadliest malignancy in women.
- In 2020, around 2.3 million women were diagnosed with breast cancer and around 685,000 deaths were recorded worldwide due to this malignancy.
- It may strike pubescent women from all over the world at any time,
- The risk increases with age.

The frequency of breast cancer continues to increase [2].

The exact causes of breast cancer are unknown. They are complex, multifactorial and depend on personal characteristics [3]. The study [4] reported that environmental factors are most likely the primary reasons behind breast cancer rather than genetic factors. According to [5], the causes of breast cancer were believed to be either behavioral (e.g. physical inactivity and alcohol intake) or non-behavioral (age and genetic factors).

However, the most common causes of breast cancer were found to be: (i) stress and worry, (ii) diet and eating habits, (iii) altered immunity, (iv) overwork and, (iv) poor previous medical care. Similarly, [3] found that stress and lifestyle-associated factors (e.g. diet, exercise and weight control) are the predominant causal agents of breast cancer.

Different therapies are used to treat breast cancer. This depends on cancer subtype and progression [6]. For example, surgery, with either lumpectomy (i.e. removing the tumor) or total mastectomy (i.e. removing the whole breast); radiotherapy, which includes external radiotherapy (i.e. high energy radiations are aimed at the breast tumor) and brachytherapy (i.e. a radioactive compound is injected into the breast); chemotherapy, through drug administration to prevent metastasis or reduce the size of an existing tumor; hormone therapy, by using either drugs, radiation or surgery to reduce female hormone activity and production; and targeted therapy coupled with immunotherapy, in which drugs are used to stimulate the immune system [7].

In Morocco, breast cancer represents around 35% of all cancers diagnosed in women, with the highest incidence rate observed in 45-59-year-old women [8] [9]. Like many other developing countries, breast cancer represents a major health concern in Morocco [10]. This is due not only to the lack of early diagnosis of this cancer, particularly in the countryside, but also to the lack of well-established treatment options for patients depending on their personal, metabolic and medical profiles. On the other hand, the acquisition of anticancer drugs is a serious problem for developing countries. Indeed, the cost of anticancer drugs used in different therapeutic regimes is extremely high and continues to rise [11]. Taking this into account along with the fact that not all Moroccan citizens and cancer treatments are covered by health insurance, and the lack of human and medical resources make it necessary to carefully manage finite resources and to optimize cancer treatment for each patient.

Today, developing countries are facing the challenge of improving patient outcomes and survival with limited access to advanced medical care [12]. Most of these countries can only provide conventional chemotherapeutic agents to their citizens [13]. Indeed, targeted therapies such as cyclin-dependent kinase inhibitors, anti-human epidermal growth factor receptor

2 (anti-HER2), tyrosine kinase inhibitors and bifunctional monoclonal antibodies (i.e. antibody–drug conjugates) are available in limited quantities and not covered by health insurance [14] [15]. Along this line, determining the optimal therapy depending on each patient specific characteristics would improve patient outcomes and satisfaction even when resources are limited [16]. In such context, the use of medical datasets would be of great interest to implement a convenient decision-making program to help physicians in identifying the most appropriate therapy for a particular patient, depending on her personal characteristics and hospital data.

In many fields, the use of datasets contributed to develop effective solutions for problematic situations. There are different types of datasets that can be deployed. For example, online datasets, generally created by experts, public or private organizations and that are accessible (for free or upon payment) on their websites; machine-generated datasets to help users to solve specific problems or to retrieve specific data; and datasets related to a specific profession or organization, and that contain contextual data based on their history. Over the years, the latter type of datasets has been used in various fields to analyze actual data and obtain valuable knowledge to handle particular situations. In the medical and health-care field, clinical-administrative datasets would allow for capitalization, management and retrieval of relevant information [17].

The objective of this work is to create and analyze a real dataset on the clinical and personal characteristics (e.g. patient age, cancer stage at diagnosis, cancer subtype) of 601 patients with breast cancer and hospitalized, and to identify the different chemotherapeutic regimens used. The findings of this study would help to: (i) define the most appropriate breast cancer treatment depending on the individual characteristics of patients; (ii) address the specific needs of each patient; (iii) facilitate the design of personalized treatment for breast cancer patients; (iv) develop a predictive algorithm to forecast the stock of chemotherapeutic molecules; and (v) better use the limited resources. The key is to have a database that can be used to realize these ideas. To the best of our knowledge the created dataset and this study has not been proposed before.

II. MATERIALS AND METHODS

A. Study Context and Sampling

In Morocco, the health care system consists of three different sectors: private non-profit, private for-profit (i.e. private clinics) and public sectors (i.e. government hospitals). Within the public sector, University Hospital Centers (UHCs) are the best-equipped facilities in terms of bed capacity and medical equipment. They include several medical disciplines and are involved in medical training and research. Thus, many Moroccan patients tend to go to UHCs for diagnosis and treatment. Based on the above, Mohammed VI UHC was chosen to perform the following investigation. The study was conducted from 2018, and included all the patients that were admitted to the Oncology and Hematology Center (OHC) of Mohammed VI UHC for breast cancer chemotherapy.

B. Data Collection

The data collection procedure is summarized in Fig. 1. Patients admitted to the OHC were categorized according to

cancer type. Only breast cancer patients to whom chemotherapy was recommended were included in this study. Cancer stage was not an inclusion/exclusion criterium. However, patients who did not receive previous adjuvant or palliative chemotherapy were not included. Based on the European Society for Medical Oncology (ESMO) clinical practice guidelines for breast cancer treatment [18], the National Comprehensive Cancer Network (NCCN) guidelines [19], and at the medical oncologist's discretion, patients with the following criteria were not included:

- Early-stage luminal A breast cancer that has not spread to lymph nodes or affected 1-3 axillary nodes, with low clinical risk of recurrence, including tumor size and histological grade.
- Node negative pT1a, triple-negative breast cancers < 5mm in diameter.
- Node negative pT1b, HER2-positive or triple negative breast cancers.
- Patients with hormone receptors positive, HER2-negative and metastatic breast cancer who received endocrine therapy alone during the inclusion period.

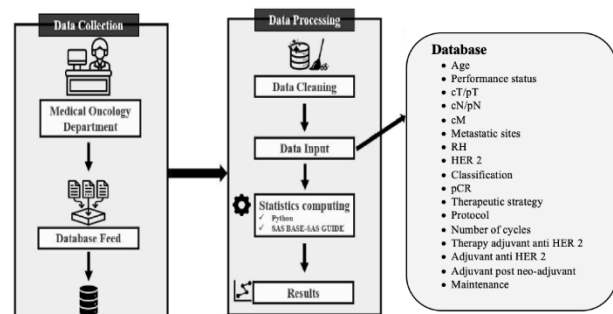


Fig. 1. Flow scheme of the study.

Data were recorded for patient demographics, breast cancer-related information including clinical and histopathological features, treatment options for patients and determinant factors that may affect or facilitate the decision-making process on whether or not to start a chemotherapy regimen, regardless if it is in the neoadjuvant, adjuvant or palliative setting. For each patient, therapy details were collected (i.e. schedule, number of cycles and the chemical dose administered). The performance status (PS) of each patient was evaluated and noted from 0 to 3 according to the Eastern Cooperative Oncology Group (ECOG) scale, with 0 being fully functional patient and 3 refers to bedridden patients. Tumor staging was assessed and classified according the American Joint Committee on Cancer (AJCC) classification system.

C. Data Analysis

Several languages including Python Program Language (PPL) were used to analyze, mine and synthesize the dataset. Statistical Analysis System (SAS) was used to generate correlation coefficients among variables and conduct a

principal component analysis (PCA) on the collected data. Prior to SAS analysis, all qualitative variables were converted into quantitative variables. Matplotlib and Seaborn were used for data visualization.

D. Ethical Considerations

Data were provided by the management of the OHC of Mohammed VI UHC following the approval of the head of the Medical Oncology Department. During our investigation, the identity of patients was never revealed.

III. RESULTS

From January, a total of 739 breast cancer patients were admitted to the OHC of Mohammed VI UHC. This accounted for 25% of the total new cancer cases. Based on the inclusion criteria, 601 patients were selected for assessment. The mean age of the patients was 48.84 years (standard deviation, 11.12; median, 48.74 years; range, 24-85 years). At the time of diagnosis, 240 patients were premenopausal, which accounted for 39.93% of the total patients. Analysis of the data showed that most patients (57.40 %) were diagnosed in locally advanced stage or disseminated stages of disease and thus required systemic chemotherapy as the mainstay of treatment. All patients had a PS score of 0 or 1 on the ECOG scale. Patients with such scores are able to withstand chemotherapy. Indeed, clinical use of chemotherapy should be restricted to medically fit patients with the best PS scores and able to tolerate aggressive therapeutic regimens.

Based on our analysis, four main subtype profiles were found: hormone receptor positive and HER2 negative (54.47 %, n = 329), hormone receptor positive or negative and HER2 positive (25.29%, n = 152), and triple negative (TNBC) (19.80 %, n = 119). Pathologic complete response (pCR) was achieved in 23 patients (3.82%). All the patients included in the dataset received chemotherapy in either the adjuvant setting (67.22 %, n = 404), neo-adjuvant setting (18.63%, n= 112), adjuvant/post-neoadjuvant setting (1.99 %, n= 12) or in palliative intent (11.64 %, n = 70) (Fig. 2).

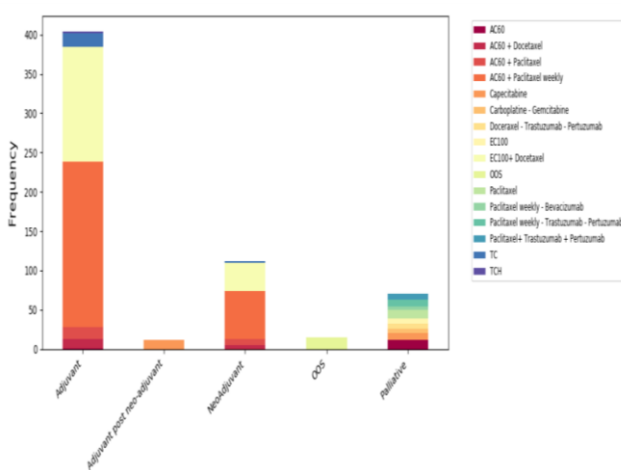


Fig. 2. Therapeutic strategy according to the prescribed treatments.

The distribution of patients according to their age was normal (Fig. 3).

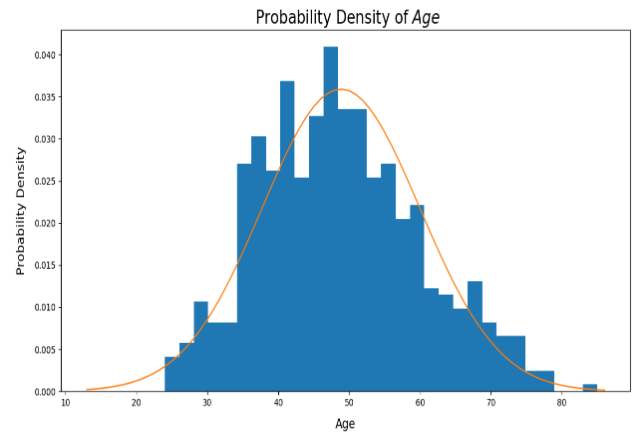


Fig. 3. Distribution of the variable age.

The distribution of patient age as a function of treatment regimen was also normal (Fig. 4). Sequential anthracycline and taxane-based regimen was the most prescribed therapy of curative intent (Fig. 4). Based on our findings, patient age appears to be a determinant factor for therapeutic decisions. Overall, the median number of pre- and postoperative chemotherapy cycles was 6, with a range of 4-8.

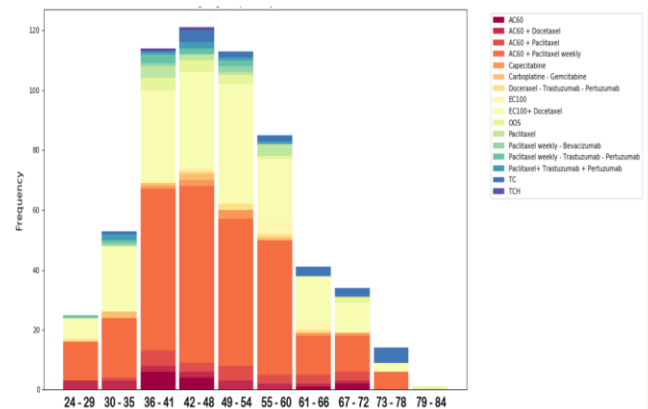


Fig. 4. Prescribed treatments according to patient age.

The subset of patients as defined by the AJCC was as follows: stage I (7.32 %, n = 44), stage IIA (18.30 %, n = 110), stage IIB (13.31 %, n = 80), stage IIIA (22.96 %, n = 138), stage IIIB (11.98 %, n = 72), stage IIIC (10.65 %, n = 64), and stage IV (11.81 %, n = 71). On the other hand, the cancer development stage of 22 patients (3.66%) could not be classified according to the AJCC criteria due to diagnostic failure or lack of information. As expected, the subtype classification of breast cancer was not correlated with anatomical prognostic factors, but can predict the tumor biological behavior.

The prescribed treatments for patients with first-line stage IV breast cancer were as follows: AC 60 was prescribed for 12 patients (17.14 %); Paclitaxel, Trastuzumab and Pertuzumab were prescribed for 11 patients (15.71 %); Paclitaxel, Pertuzumab and weekly Trastuzumab were prescribed for 9 patients (12.86 %); Paclitaxel and weekly-Bevacizumab were prescribed for 8 patients (11.43 %); EC100 was prescribed for

7 patients (10 %); Paclitaxel was prescribed for 7 patients (10%); Carboplatine and Gemcitabine were prescribed for 6 patients (8.57 %); Doceraxel, Trastuzumab and Pertuzumab were prescribed for 6 patients (8.57%); and Capecitabine was prescribed for 4 patients (5.71 %) (Fig. 2).

Fig. 5 shows the results of correlation among the different variables. According to practicing physicians, the variables presented in Fig. 5 are the most important ones for treatment prescription.

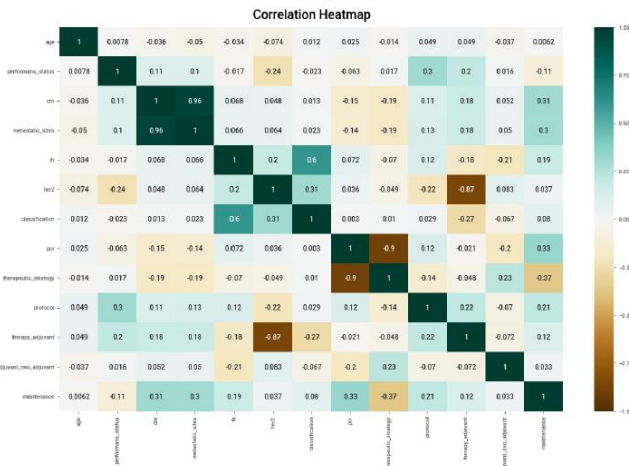


Fig. 5. Principal component analysis of causal attributions of breast cancer.

The correlation circle (Fig. 6) shows the distance between variables. Based on Fig. 5 and 6, it could be concluded that maintenance, therapeutic strategy, metastatic sites, cM and cT/pT were the main variables positively correlated with component 1 while HER2 and therapy adjuvant anti-HER2 were the main variables positively correlated with component 2. On the other hand, classification was the main variable negatively correlated with component 2. The findings of this study would be very useful in developing a machine learning model for the prediction of breast cancer treatment under Moroccan circumstances.

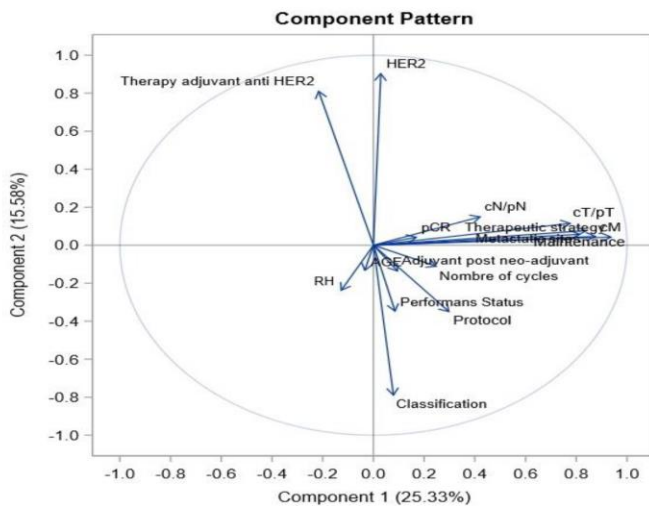


Fig. 6. Correlation circle of causal attributions of breast cancer.

IV. DISCUSSION

Breast cancer is one of the most common cancers diagnosed in women. This disease causes a devastatingly high number of deaths every year [20]. In developing countries, controlling breast cancer is more problematic due to the lack of medical and human resources [21]. Furthermore, patients in developing countries have limited access to early diagnosis and to modern diagnostic tools [22]. In such circumstances, developing treatment prediction strategies could substantially contribute to save thousands of lives each year, especially when the tumor is detected in an advanced stage. The present study was conducted to analyze personal and medical data of breast cancer patients in Morocco. This would help to develop a predictive algorithm to determine the most appropriate treatment for patients, to better use the limited resources and to accurately estimate hospital needs.

In recent years, the use of datasets to assist in the decision making process has known a great success. This was observed in different fields including the medical one (e.g. [23], [24], [25]). Indeed, in the medical and health-care field, datasets are increasingly involved in planning and managing limited resources and to assist the decision-making process. According to [26], medical datasets are highly relevant for cancer research and clinical application. Medical datasets are composed of records collected during the clinical care process, and generally include individual records of patients, medical diagnoses, drug prescription and treatments [27]. In Morocco, the use of datasets to manage resources and take decisions is still in its infancy. Collecting, analyzing and processing breast cancer patient data would help in better utilizing the limited human and medical resources, and could improve the decision-making process depending on personal and medical records of patients.

The findings of this investigation showed that most breast cancer patients admitted to Mohamed VI UHC were 24-85-year-old, with a mean age of 48.84 years. Population-based data (271,173 patients) from Surveillance, Epidemiology, and End Results (SEER) cancer registries showed that more than 94% of breast cancer patients registered in 18 different USA cities/states and diagnosed between 2010 and 2015 were over 40 years old, with a mean age of 60 years [28]. Similarly, [25] found that among 235,368 French patients diagnosed with early breast cancer between 2011 and 2017 (identified by the Oncology Data Platform -(ODP) of the French National Cancer Institute (INCa)), 95% were 40 years old or above at the time of diagnosis. [29] used the Queensland Cancer Registry and found that among 3,079 women diagnosed with breast cancer in Queensland (Australia) between July 1st 2011 and June 30th 2012, 87.9% were older than 45 years. Based on our findings and the above-mentioned reports from the literature, it could be assumed that most women diagnosed with breast cancer are over 40 years of age. Up to date, the mechanisms underlying the relationship between age and breast cancer incidence remains unclear [30]. Indeed, the factors that induce breast cancer are complex, multiple and vary depending on the disease subtype [31][32]. According to many studies, age is not a determining factor that influences the incidence of breast cancer [33]. On the other hand, other studies indicated that in young patients, the pathophysiology of breast cancer is attributed to pathways linked to phosphoinositide 3-kinase,

immature mammary epithelial cells, growth factor signaling and mitogen activated protein kinase [34]. These diverging interpretations underline that research findings are not sufficiently conclusive regarding the likely effect of age on breast cancer incidence.

The findings of the present study showed that most women (45.59 %) had stage III breast cancer (based on the AJCC system), which is considered as locally advanced stage. Indeed, stage III is the stage where the tumor size is larger than 5 cm with involved lymph nodes [35]. In developing countries, most patients are diagnosed at advanced stages, which results in poor response to treatment. Examples from the literature showed that in developed countries most breast cancer cases are diagnosed at earlier stage. For example, [36] found that stage I tumors were the most common among 7,458 patients diagnosed at Asan Medical Center (Seloul, Korea) between January 1999 and December 2008. The researcher in [37] used data from 18 cancer registries (18 US cities/states) of the SEER program to determine the stage of breast cancer in 14,379 young women between 2010 and 2014. They found that most patients had tumors at stages I and II (28.2% and 45.2%, respectively), while only 19.0% had stage III breast cancer at diagnosis. The study [38] reported that among 1,703 breast cancer women diagnosed at the McGill University Health Centre (Montreal, Canada) between January 2010 and December 2013, the majority had stage I cancer. Stage I breast cancer is the stage where the tumor size is no more than 2 cm, with no involved axillary lymph nodes and without metastatic disease [39]. The fact that in Morocco most breast cancer patients are diagnosed at stage III raises the concern of late diagnosis, which would affect the efficiency of treatments and recovery rates. Taking this into account along with the limited resources in developing countries, developing efficient methods to determine the most effective treatment for each patient seems crucial.

Our findings showed that 11.81% of patients had a stage IV tumor. Stage IV cancer is that when the tumor has spread to other organs of the body (i.e. distant metastatic disease) [39]. Data from Mohamed VI UHC revealed that at this stage, nine different treatments were prescribed, with percentages varying from 5.71 to 17.14 %. This large number of treatments with close percentages of prescription may reflect the challenge of identifying the most effective treatment for stage IV breast cancer and highlights the importance of using datasets to decide on the most appropriate treatment depending on patient characteristics. Our findings also revealed that most patients have received systemic chemotherapy as the mainstay of treatment. Some currently well-established treatment methods that include escalating, de-escalating or omitting cytotoxic chemotherapy are not applicable in our context due to limited or no access to precision medicine. For example, the currently available gene expression signatures used to assess the risk of recurrence and the benefit of adjuvant chemotherapy in intermediate risk luminal breast cancers are not freely available in Morocco. Therefore, physicians rely solely on clinical-pathological factors, which lead to over-prescription of chemotherapy. Correlations among the variables included in the dataset confirm the fact that the treatment regimen depends

mainly on cancer stage, which relies on clinical-pathological features.

Our study showed that the most prescribed treatment was the combination of Paclitaxel, Trastuzumab and Pertuzumab. Paclitaxel is a natural anticancer agent widely used to treat cancer patients due to its particular mechanism of action and successful outcomes [40]. Trastuzumab is a monoclonal antibody generally used to treat some HER2-positive tumors such as metastatic gastric cancer and breast cancer [41] [42][43]. This agent is known for its high cost, and in many countries it is not covered by health insurance. Consequently, in many countries patients cannot afford it [41]. Regarding Pertuzumab, it is an HER dimerization inhibitor and a major drug used to treat early and HER2-positive metastatic breast cancers [44][45]. The research [46] examined data from the Ontario Cancer Registry between April 1st 2012 and March 31st 2016 and found that, among 190 triple-negative breast cancer patients at stage IV, 25.3% have underwent surgery, 72.6% received systemic therapy and 58.9% received radiotherapy. They also indicated that the top drug regimens include anthracyclines and/or taxanes. On the other hand, these authors found that the cost of treatment for stage IV patients was four times higher than that of stage I-III patients. Based on the National Cancer Dataset (NCDB) data from 2010 to 2015, [47] compared the effects of different treatments on stage IV breast cancer patients from the US (i.e. 12,838 women who lived longer than six months after their diagnosis) and found that the most effective treatments were based on the combination of either systemic therapy (i.e. chemotherapy, endocrine therapy, or both) and surgery, or systemic therapy, surgery and radiation as compared to systemic therapy alone. These findings from the literature showed that the treatment for stage IV breast cancer may vary from systemic therapy, radiotherapy, surgery or their combination. In Morocco, while only conventional chemotherapeutic agents are either freely available or are covered by health insurance, the current challenges are to determine the most appropriate treatment for each patient and to avoid that demand outweighs supply. The use of datasets would considerably help in determining the most effective treatment depending on patient characteristics and tumor phenotype.

V. CONCLUSION

The findings of the present study provided valuable data on breast cancer patients in Morocco and the treatment regimens used. Similarly to many other developing countries, breast cancer is generally diagnosed at an advanced stage (i.e. stage III and stage IV) in Morocco. Systemic chemotherapy was found to be the mainstay of treatment. Nine different treatments were prescribed. The chemotherapeutic regimens used reflect the lack of a single standard or optimized treatments for patients. Further research could focus on evaluating the outcome of the different treatments, and to develop a predictive algorithm to determine the most appropriate treatment for each patient depending on demographic and medical characteristics. This would help to rationally use the limited medical and human resources and to improve patient outcomes and satisfaction.

VI. DECLARATIONS

A. Funding

This study did not receive any support from external funding agencies.

B. Conflict of Interest

The authors declare they have no conflict of interest.

C. Data Availability

Data available upon request from the corresponding author.

D. Ethics Approval and Consent to Participate

Human Research Ethics approval was obtained from the Mohamed VI UHC, Medical Oncology Department.

REFERENCES

- [1] WHO (2021) World Health Organization. Breast Cancer. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Scope%20of%20the%20problem,the%20world's%20most%20prevalent%20cancer.> Accessed Jun 08th 2022.
- [2] Susini T, Biglia N, Bounous VE (2022) Prognostic factors research in breast cancer patients: new paths. *Cancers* 14:971. <https://doi.org/10.3390/cancers14040971>.
- [3] Park SK, Min YH, Lee M, Lee SB (2021) Causal attributions in breast cancer patients planning to undergo adjuvant endocrine therapy. *Int J Environ Res Public Health* 18:5931. <https://doi.org/10.3390/ijerph18115931>.
- [4] Bessonau V, Rudel RA (2020) Mapping the human exposome to uncover the causes of breast cancer. *Int. J. Environ. Res. Public Health* 17:189. <https://doi.org/10.3390/ijerph17010189>.
- [5] Lee Y, Jeon Y-W, Im E-O, Baek J-M (2021) Causal attributions and quality of life of Korean breast cancer survivors. *Asian Nurs Res* 15:53–59. <https://doi.org/10.1016/j.anr.2020.11.002>.
- [6] Seo J, Guk G, Park S-H, Jeong M-H, Jeong J-H, Yoon H-G, Choi K-C (2019) Tyrosine phosphorylation of HDAC3 by Src kinase mediates proliferation of HER2-positive breast cancer cells. *J Cell Physiol* 234:6428–6436. <https://doi.org/10.1002/jcp.27378>.
- [7] Alphan ery E (2014) Perspectives of breast cancer thermotherapies. *J Cancer* 5:472–479. <https://doi.org/10.7150/jca.8693>.
- [8] Khalis M, El Rhazi K, Charaka H, Chaj es V, Rinaldi S, Nejari C, Romieu I, Charbotel B (2016) Female breast cancer incidence and mortality in Morocco: comparison with other countries. *Asian Pac J Cancer Prev* 17:5211–5216. <https://doi.org/10.22034/APJCP.2016.17.12.5211>.
- [9] Khalis M, El Rhazi K, Fort E, Chaj es V, Charaka H, Huybrechts I, Moskal A, Biessy C, Romieu I, Abbass F, El Marnissi B, Mellas N, Nejari C, Soliman AS, Charbotel B (2019) Occupation and risk of female breast cancer: a case-control study in Morocco. *Am J Ind Med* 62:838–846. <https://doi.org/10.1002/ajim.23027>.
- [10] Charaka H, Khalis M, Elfakir S, Huybrechts I, Chami Khazraji Y, Lyoussi B, Soliman AS, Nejari C (2021) Knowledge, perceptions, and satisfaction of Moroccan women towards a new breast cancer screening program in Morocco. *J Canc Educ* 36:657–663. <https://doi.org/10.1007/s13187-019-01680-6>.
- [11] Yamada H, Kobayashi R, Shimizu S, Yamada Y, Ishida M, Shimoda H, Kato-Hayashi H, Fujii H, Iihara H, Tanaka H, Suzuki A (2020) Implementation of a standardised pharmacist check of medical orders prior to preparation of anticancer drugs to reduce drug wastage. *Int J Clin Pract* 74:e13464 <https://doi.org/10.1111/ijcp.13464>.
- [12] Chen G, Xiao X, Zhao X, Tat T, Bick M, Chen J (2021) Electronic textiles for wearable point-of-care systems. *Chem Rev* 122:3259–3291. <https://doi.org/10.1021/acs.chemrev.1c00502>.
- [13] Chen D, Si W, Shen J, Du C, Lou W, Bao C, Zheng H, Pan J, Zhong G, Xu L, Fu P, Fan W (2018) miR-27b-3p inhibits proliferation and potentially reverses multi-chemoresistance by targeting CBLB/GRB2 in breast cancer cells. *Cell Death Dis* 9:188. <https://doi.org/10.1038/s41419-017-0211-4>.
- [14] Li J, Wang S, Wang Y, Wang X, Wang H, Feng J, Zhang O, Sun T, Ouyang O, Yin Y, Liu Y, Geng C, Yan M, Jiang Z (2017) Disparities of trastuzumab use in resource-limited or resource-abundant regions and its survival benefit on HER2 positive breast cancer: a real-world study from China. *Oncologist* 22:1333–1338. <https://doi.org/10.1634/theoncologist.2017-0088>.
- [15] Chen Z, Ouyang O, Wang Y, Wang J, Wang H, Wu X, Zhang P, Huang J, Zheng Y, Cao W, Shao X, Xie N, Tian C, Liang H, Wang C, Zhang Y, Ren D, Wang X (2022) Real-world first-line treatment patterns and outcomes in hormone receptor-positive advanced breast cancer patients: a multicenter, retrospective study in China. *Front Oncol* 12:829693. <https://doi.org/10.3389/fonc.2022.829693>.
- [16] Vokinger KN, Hwang TJ, Grischoff T, Reichert S, Tibau A, Rosemann T, Kesselheim AS (2020) Prices and clinical benefit of cancer drugs in the USA and Europe: a cost–benefit analysis. *Lancet Oncol* 21:664–670. [https://doi.org/10.1016/S1470-2045\(20\)30139-X](https://doi.org/10.1016/S1470-2045(20)30139-X).
- [17] Owens M-R, Nguyen S, Karsy M (2022) Utility of administrative datasets and big data on understanding glioma treatment-a systematic review. *Indian J Neurosurg*. <https://doi.org/10.1055/s-0042-1742333>.
- [18] Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, Zackrisson S, Senkus E (2019) Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 30:1194–1220. <https://doi.org/10.1093/annonc/mdz173>.
- [19] NCCN (2021) Treatment by Cancer Type. NCCN Guidelines. https://www.nccn.org/guidelines/category_1. Accessed Jun 14, 2022.
- [20] Raj S, Singh S, Kumar A, Sakar S, Pradhan C (2021) Feature selection and random forest classification for breast cancer disease. In: Satpathy R, Choudhury T, Satpathy S, Mohanty SN, Zhang X (eds) *Data analytics in bioinformatics*. Wiley: Hoboken, NJ, USA, pp 191–210. <https://doi.org/10.1002/9781119785620.ch8>.
- [21] Zhuang Q, Wu J, Yu G (2021) A medical support system for prostate cancer based on ensemble method in developing countries. In: He X, Shao E, Tan G (eds) *Network and parallel computing. NPC 2020. Lecture Notes in Computer Science*, vol. 12639. Springer, Cham, pp 361–372. https://doi.org/10.1007/978-3-030-79478-1_31.
- [22]  akmak GK, Emiro lu S, Sezer A, Canturk NZ, Yeniay L, Kuru B, Karanlık H, Soyder A, Gokgoz S, Sakman G, Ucuncu M, Akcay MN, Girgin S, Gurdal SO, Emiroglu M, Ozbas S, Bahadir A, Arici C, Toktas O, Demirean O,  alik A, Polat AK, Maralean G, Demire D, Ozmen. V (2020) Surgical trends in breast cancer in Turkey: an increase in breast-conserving surgery. *JCO Glob Oncol* 6:285–292. <https://doi.org/10.1200/JGO.19.00275>.
- [23] Mazo C, Kearns C, Mooney C, Gallagher WM (2020) Clinical decision support systems in breast cancer: A systematic review. *Cancers* 12:369. <https://doi.org/10.3390/cancers12020369>.
- [24] Agossou C, Atchad  MN, Djibril AM, Kurisheva SV (2022) Mathematical modeling and machine learning for public health decision-making: the case of breast cancer in Benin. *Math Biosci Eng* 19:1697–1720. <https://doi.org/10.3934/mbe.2022080>.
- [25] Dumas E, Laot L, Coussy F, Grandal Rejo B, Daoud E, Laas E, Kassara A, Majdling A, Kabirian R, Jochum F, Gougis P, Michel S, Houzard S, Le Bihan-Benjamin C, Bousquet P-J, Hotton J, Azencott C-A, Reyat F, Hamy A-S (2022) The French Early Breast Cancer Cohort (FRESH): a resource for breast cancer research and evaluations of oncology practices based on the French National Healthcare System Dataset (SNDS). *Cancers* 14:2671. <https://doi.org/10.3390/cancers14112671>.
- [26] Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN, Collins HL, Nordyke RJ, Whyte JL (2011) Use of electronic medical records (EMR) for oncology outcomes research: Assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol* 3:259–272. <https://doi.org/10.2147/CLEP.S23690>.
- [27] Hennessy S (2006) Use of health care datasets in pharmacoepidemiology. *Basic Clin Pharmacol Toxicol* 98:311–313. https://doi.org/10.1111/j.1742-7843.2006.pto_368.x.
- [28] Kim HJ, Kim S, Freedman RA, Partridge AH (2022) The impact of young age at diagnosis (age <40 years) on prognosis varies by breast

- cancer subtype: A U.S. SEER dataset analysis. *Breast* 61:77–83. <https://doi.org/10.1016/j.breast.2021.12.006>.
- [29] Bates N, Callander E, Lindsay D, Watt K (2020) Patient co-payments for women diagnosed with breast cancer in Australia. *Support Care Cancer* 28:2217–2227. <https://doi.org/10.1007/s00520-019-05037-z>.
- [30] Rozenblit M, Hofstatter E, Liu Z, O'Meara T, Stormiolo AM, Dalela D, Singh V, Pusztai L, Levine M (2022) Evidence of accelerated epigenetic aging of breast tissues in patients with breast cancer is driven by CpGs associated with polycomb-related genes. *Clin Epigenet* 14:30. <https://doi.org/10.1186/s13148-022-01249-z>.
- [31] Calaf GM, Ponce-Cusi R, Aguayo F, Muñoz JP, Bleak TC (2020) Endocrine disruptors from the environment affecting breast cancer. *Oncol Lett* 20:19–32. <https://doi.org/10.3892/ol.2020.11566>.
- [32] Elaraby E, Malek AI, Abdullah HW, Elemam NM, Saber-Ayad M, Talaat IM (2021) Natural killer cell dysfunction in obese patients with breast cancer: A review of a triad and its implications. *J Immunol Res* 2021:9972927. <https://doi.org/10.1155/2021/9972927>.
- [33] Wong FY, Tham WY, Nei WL, Lim C, Miao H (2018) Age exerts a continuous effect in the outcomes of Asian breast cancer patients treated with breast-conserving therapy. *Cancer Commun* 38:39. <https://doi.org/10.1186/s40880-018-0310-3>.
- [34] Shin H-C, Han W, Moon H-G, Im S-A, Moon WK, Park I-A, Park SJ, Noh D-Y (2013) Breast-conserving surgery after tumor downstaging by neoadjuvant chemotherapy is oncologically safe for stage III breast cancer patients. *Ann Surg Oncol* 20:2582–2589. <https://doi.org/10.1245/s10434-013-2909-6>.
- [35] Lee SB, Sohn G, Kim J, Chung IY, Lee JW, Kim HJ, Ko BS, Son BH, Ahn SH (2018) A retrospective prognostic evaluation analysis using the 8th edition of the American Joint Committee on Cancer staging system for breast cancer. *Breast Cancer Res Treat* 169:257–266. <https://doi.org/10.1007/s10549-018-4682-5>.
- [36] Franzoi MA, Schwartzmann G, de Azevedo SJ, Geib G, Zaffaroni F, Liedke PER (2019) Differences in breast cancer stage at diagnosis by ethnicity, insurance status, and family income in young women in the USA. *J Racial Ethn Health Disparities* 6:909–916. <https://doi.org/10.1007/s40615-019-00591-y>.
- [37] Savage P, Yu N, Dumitra S, Meterissian S (2019) The effect of the American Joint Committee on Cancer eighth edition on breast cancer staging and prognostication. *Eur J Surg Oncol* 45:1817–1820. <https://doi.org/10.1016/j.ejso.2019.03.027>.
- [38] Russell CA (2003) Adjuvant systemic therapy for lymph node-negative breast cancer less than or equal to 1 cm. *Curr Oncol Rep* 5:72–77. <https://doi.org/10.1007/s11912-003-0090-y>.
- [39] Tian X, Yu H, Li D, Jin G, Dai S, Gong P, Kong C, Wang X (2021) The mir-5694/Af9/Snai1 axis provides metastatic advantages and a therapeutic target in basal-like breast cancer. *Mol Ther* 29:1239–1257. <https://doi.org/10.1016/j.ymthe.2020.11.022>.
- [40] Zhu L, Chen L (2019) Progress in research on paclitaxel and tumor immunotherapy. *Cell Mol Biol Lett* 24:40. <https://doi.org/10.1186/s11658-019-0164-y>.
- [41] Sarosiek T, Morawski P (2018) Trastuzumab and its biosimilars. *Polski Merkuriusz Lekarski* 44:253–257.
- [42] Akbari V, Chou CP, Abedi D (2020) New insights into affinity proteins for HER2-targeted therapy: beyond trastuzumab. *Biochim Biophys Acta Rev Cancer* 1874:188448. <https://doi.org/10.1016/j.bbcan.2020.188448>.
- [43] Waller CF, Möbius J, Fuentes-Alburo A (2021) Intravenous and subcutaneous formulations of trastuzumab, and trastuzumab biosimilars: implications for clinical practice. *Br J Cancer* 124:1346–1352. <https://doi.org/10.1038/s41416-020-01255-z>.
- [44] Robert M, Frenel JS, Bourbouloux E, Rigaud DB, Patsouris A, Augereau P, Gourmelon C, Campone M (2020) Pertuzumab for the treatment of breast cancer. *Expert Rev Anticancer Ther* 20:85–95. <https://doi.org/10.1080/14737140.2019.1596805>.
- [45] Jagosky M, Tan AR (2021) Combination of pertuzumab and trastuzumab in the treatment of HER2-positive early breast cancer: A review of the emerging clinical data. *Breast Cancer* 13:393–407. <https://doi.org/10.2147/BCTT.S176514>.
- [46] Brezden-Masley C, Fathers KE, Coombes ME, Pourmirza B, Xue C, Jerzak KJ (2020) A population-based comparison of treatment patterns, resource utilization, and costs by cancer stage for Ontario patients with triple-negative breast cancer. *Cancer Med* 9:7548–7557. <https://doi.org/10.1002/cam4.3038>.
- [47] Stahl K, Wong W, Dodge D, Brooks A, McLaughlin C, Olecki E, Lewcun J, Newport K, Vasekar M, Shen C (2021) Benefits of surgical treatment of stage iv breast cancer for patients with known hormone receptor and HER2 status. *Ann Surg Oncol* 28:2646–2658. <https://doi.org/10.1245/s10434-020-09244-5>.

Presenting a Novel Method for Identifying Communities in Social Networks Based on the Clustering Coefficient

Zhihong HE^{1*}, Tao LIU²

College of Art and Design, Chongqing Vocational College of Culture and Arts, Chongqing 400067, China¹
College of Artificial Intelligence and Big Data, Chongqing College of Electronic Engineering, Chongqing 401331, China²

Abstract—In recent decades, social networks have been considered as one of the most important topics in computer science and social science. Identifying different communities and groups in these networks is very important because this information can be useful in analyzing and predicting various behaviors and phenomena, including the spread of information and social influence. One of the most important challenges in social network analysis is identifying communities. A community is a collection of people or organizations that are more densely connected than other network entities. In this article, a method to increase the accuracy, quality, and speed of community detection using the Fire Butterfly algorithm is presented, which defines the algorithm and fully introduces the parameters used in the proposed algorithm and how to implement it. In this method, first the social network is converted into a graph and then the clustering coefficient is calculated for each node. Also, the butterfly algorithm based on the clustering coefficient (CC-BF) has been proposed to identify complex social networks. The proposed algorithm is new both in terms of generating the initial population and in terms of the mutation method, and these improve its efficiency and accuracy. This research is inspired by the meta-heuristic algorithm of Butterfly Flame based on the clustering coefficient to find active nodes in the social network. The results have shown that the proposed algorithm has improved by 23.6% compared to previous similar works. The findings of this research have great value and can be useful for researchers in computer science, social network managers, data analysts, organizations and companies, and other general public.

Keywords—Social network; detection of communities; butterfly fire algorithm; clustering coefficient

I. INTRODUCTION

Today, the Internet and web services are expanding rapidly, and at the same time, virtual social networks play an important role in people's real lives [1]. In fact, social networks are interactive networks that use the Internet as a medium to create communication between people [2]. With the rapid increase in social network users, high-scale data exploration can provide a better and more effective view of the hidden potential of these networks [3]. Internet social networks, as the most important examples, for the presence of different segments of society and the exchange of ideas, thoughts, and needs have changed according to social life [4]. A social network is a social structure formed by a group of people, organizations, or other social entities [5]. This collection is connected by social relations such as information exchange, cooperation,

friendship, kinship, or financial exchanges [6]. Social network analysis is an approach used to study the interaction of humans and examine the patterns, communication structure, or organization of social networks [7]. With the expansion of the use of electronic and online communication methods, the number of social networks has increased, and the importance of extracting communities from these networks in order to analyze social networks has become more important [8]. In social networks, some nodes are more connected than the entire network's nodes, which are called communities. Nodes are connected by one or more specific types of dependencies [9]. For example, financial exchanges, friendships, kinship, business, web links, disease transmission, or airline routes are examples of communication. But the resulting structure of these networks is often very extensive and complex. Analysis of social networks is the mapping and measurement of cooperation relationships among individuals, groups, organizations, and any entity that has the ability to process information and knowledge. Graph theory is usually used to display and analyze social networks. The components of graph theory are nodes and edges [10].

Groups of nodes (or members) connected by one or more different kinds of relationships are referred to as social networks [11]. In social network analysis, network structure is understood as the pattern organization of those nodes and their relationships, which helps to explain how these patterns have an impact on people's behaviors and attitudes [12]. Different communities inside the network are formed by these exchanges, links, or connections [13]. Those communities' members frequently share certain traits or interests. Comprehending and using the network effectively begin with comprehending the structure of society [14]. In the literature, clustering and community detection are frequently used interchangeably. While community detection strategies for network analysis and an emphasis on network structure are created as a function of connectivity that incorporates social interaction, clustering techniques typically concentrate on a single strategy, such as using particular traits to group nodes in a network [15]. Clustering algorithms, on the other hand, can be thought of as a workable substitute for community detection techniques, and both of them can be used to solve a variety of network analysis problems [16]. Existing paper swarm algorithm-based methods have drawbacks, including a slow rate of community discovery, among others. With increasing fitness value, the detection speed of communities slows down,

which could lead to a local optimum and take some time to attain a virtually global optimum [17]. A near-global optimum is quickly reached when efficiency is decreased, which lowers the likelihood of becoming stuck in a local optimum like speed detection. If an algorithm has improved speed detection, it may be able to reach a near-global optimum more quickly and with fewer iterations. In order to detect communities in a network, a new solution for the butterfly-fire technique based on the clustering coefficient (CC-BF) is proposed in this research. In addition to identifying communities in highly populated networks, CC-BF discovers cohesive collections of nodes as anomalous networks. The primary concept behind CC-BF is that it leverages the clustering coefficient, a social network analysis indicator, to improve community detection's accuracy, quality, and speed.

Other academics have successfully used strategies similar to CC to locate community structures in complex networks. Existing studies, however, don't look at the application of the clustering coefficient in BF. The clustering coefficient-based butterfly-fire algorithm's (CC-BF) essential phases are as follows: The speed of community detection using the clustering coefficient Modularity-based efficiency assessment.

To evaluate the quality of the population, the modularity criterion has been applied as a fitness function. Prior information about the size, quantity, or structure of communities is not necessary for the butterfly-fire approach, which is based on the clustering coefficient. According to experimental findings, CC-BF outperforms several other approaches for various networks. The following are the paper's contributions:

- A novel mutation technique that exploits the identified community structure to rewire existing connections;
- Using the clustering coefficient for the speed of community detection using the Fire Butterfly algorithm, comparing eight cutting-edge methods and evaluating the suggested algorithm on 12 different kinds of small and large networks.

The paper is divided into the following sections: The related works are included in the second section. The third section provides the suggested solution. The evaluation and effectiveness of the suggested algorithm are examined in the fourth section, along with a comparison to other algorithms of a similar nature. Finally, the fifth section presents the conclusion.

II. RELATED WORKS

In addition to the structural analysis of social networks, identifying communities is also used in other issues such as customer classification, recommender systems, vertex labelling, analysis of network influencers, and information dissemination. Due to the complexity of this issue, despite the various efforts that have been made in this field in recent years, it has not yet been fully resolved, and a satisfactory answer has not been provided to solve this issue. Many community identification algorithms have been presented for analyzing social networks. Identifying communities can be considered an optimization problem; from this point of view, several methods

based on the maximization of the famous modularity criterion have been presented for identifying communities [18]. Among the famous methods in this field, we can mention Newman's greedy algorithm (FN) [19]. In this method, conjunctive hierarchical clustering is used, in which groups of nodes successfully form larger communities if and only if the modularity value increases after this integration. After that, Klast and his colleagues presented the CNM method [20, 34] and showed that by using a complex data structure, they were able to reduce the computational burden of modularity in Newman's algorithm, making it usable for large networks. Shang and his colleagues also presented an improved genetic algorithm called MIGA to obtain maximum modularity [21]. Fortunato and his colleagues showed in [22, 33] that this method also has limitations apart from modularity. One of the most important limitations of modularity in this field is the limitation of separability. The problem of separability has a great impact on practical problems in the real world. Because of such issues, communities have different sizes, and this problem causes many small communities not to be recognized. Two solutions have been proposed to overcome the limitation of separability: First, other quality metrics were proposed in addition to modularity to be identified at different scales of society. For example, in a paper [23], a criterion called significance was used instead of modularity as the objective function in the Levin method, and it produced better results than using modularity. Pizzotti also introduced a standard called "community rank" [24] to guarantee the high density of communication within communities and the low density of communication between communities. The second solution to this problem is to formulate community identification as a multi-objective optimization problem. Pizzotti in [25] tries to obtain suitable communities using the NSGA-II evolutionary algorithm and two objective functions of community merit and rank. Gang and his colleagues presented the MOEA/D-Net algorithm [26], which tries to optimize two objective functions against each other. In [27], the CLANet method is also presented, which uses learning automata to maximize modularity along with a local limiter. Wasserman and his colleagues [28, 32], as the pioneers of social network analysis, have cited "social" for social networks, which can be defined as a system of social relations described by a set of actors and their social connections. Usually, a social network can be shown in the form of a graph, which has a set of vertices (nodes) and edges (connections) so that vertices are considered as actors within the network and edges are considered as relationships between these actors. In its simplest form, a social network is a mapping of all the relevant edges between the studied vertices. Here are the gaps identified in previous studies and the suitability of the proposed algorithm for certain types of data:

1) *Gaps in previous studies:* Lack of Clustering Coefficient in Butterfly-Fire Algorithm (BF): The text mentions that existing studies have successfully used strategies similar to the proposed Clustering Coefficient-based Butterfly-Fire algorithm (CC-BF) for community detection. However, it notes that previous studies haven't explored the application of the clustering coefficient in the Butterfly-fire algorithm. This suggests a gap in the literature where the

combination of the clustering coefficient and the BF algorithm hasn't been extensively investigated.

Community Detection Speed: The text highlights that existing swarm algorithm-based methods for community detection have limitations, including a slow rate of community discovery. This issue could lead to suboptimal results. The proposed CC-BF algorithm aims to address this gap by improving the speed of community detection. This suggests a need for faster and more efficient community detection algorithms in the existing literature.

2) *Suitability for data types:* The text doesn't explicitly mention the types of data for which the proposed CC-BF algorithm is more suitable. However, it does mention that CC-

BF is evaluated on "12 different kinds of small and large networks." This suggests that the algorithm is designed to be versatile and applicable to a variety of network data types.

In summary, the gaps in previous studies revolve around the lack of exploration of the clustering coefficient in the context of the Butterfly-Fire algorithm and the need for faster community detection methods. The proposed CC-BF algorithm appears to be designed to address these gaps and is tested on various types of network data.

Table I shows the classification of existing studies into orderly works based on applied community detection or clustering methods.

TABLE I. CLASSIFICATION OF CASE STUDIES BASED ON COMMUNITY DETECTION METHODS OR FUNCTIONAL CLUSTERING

Tool for clustering or detecting communities	No. of studies	ID	No. of nodes (n)	Comments	Ref.
A rapid method for modularity	4	S27, s28,s45,s51	more than 5000	used to make a comparison in (S27, S45)	[16]
MapInfo algorithm	3	S,10, S13, S18	2000	used to make a comparison in (S13)	[18]
Algorithm for maximizing expectations	2	S32, S36	more than 3500	used as a comparative tool	[19]
Community detection algorithm for binary graphs	5	S16, S24, S29, S41, S47	more than 6000	used as a comparative tool	[23]
Algorithm for eigenvector label propagation	1	S03	1500	newly suggested method	[24]
Spectral clustering-based technique for left-to-right oscillation	4	S40, S46, S49, S53	more than 5000	newly suggested method	[29]
Algorithm for evaluation and identification in the community	2	S34, S45	5000	newly suggested method	[30]
Eigenvector algorithm in front	2	S22, S05	more than 4000	newly suggested method	[31]

III. SUGGESTED METHOD

The most important problem in identifying communities in social networks can be considered the speed and accuracy of identifying communities in networks, as well as the new identification methods, which, in addition to speed and accuracy, should be able to act in a way to reduce the possibility of the influence of noise on the misidentification of communities. Minimize, and with the least knowledge of the structure of networks and the number and size of associations in the network, the best number and size should be selected and identified in an excellent way. Fig. 1 shows the general design of the proposed algorithm, which is used to detect and identify active nodes in social networks. In this algorithm to detect active nodes in social networks, attention has been paid to different parts of it, which have been examined in the following:

A. Initialization

In this algorithm, it is assumed that butterflies are candidate neighbors, and the variables of the problem are the locations of the bodies in the neighborhood. Therefore, butterflies can move in a one-dimensional, two-dimensional, or multi-dimensional dimension. Since the flame-propeller optimization algorithm is a population-based algorithm, the set of propellers is a matrix of ordered $n \times d$. In this method, in the initialization stage, a random solution is generated for each moth butterfly ($k=1, 2, \dots, B$), where B represents the number of butterflies. The active node is represented by an array of length n , where

the number stored in the i index of the array shows the ID of the candidate active node that executes the task T_i . Taking into account that the butterfly will be created in the initialization step B , the initial population of solutions will be a $B \times n$ matrix.

$$M = \begin{bmatrix} m_{1,1} & \dots & m_{1,d} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \dots & m_{n,d} \end{bmatrix} \quad (1)$$

After initialization, the function P is executed iteratively until the function T is correct. The P function is the main function that moves around the search space. As mentioned above, the inspiration for this algorithm is transverse orientation. In order to mathematically model this behavior, the position of each bullet with respect to the flame is updated using the following equation.

$$M_i = S(M_i, F_j) \quad (2)$$

F_j represents the j th flame, M_i represents the i th butterfly, and s is the spiral function for the butterfly-fire algorithm.

$$S(M_i, F_j) = D_i \times e^{bt} \times \cos(2\pi t) + F_j \quad (3)$$

In this regard, D_i refers to the distance of the i -th propeller from the j -th flame. b is a fixed number to determine the shape of the logarithmic spiral, and t is a random number in $[-1,1]$, where D_i is obtained from Eq. (3).

$$D_i = |F_j - M_i| \quad (4)$$

Considering the above, the pseudocode of the proposed method is presented in Fig. 2.

In the general butterfly-fire method, all newly generated butterflies are accepted and kept in the next generation, while a greedy strategy is used to accept butterflies that have a good fit in our algorithm. This greedy strategy can be as follows:

$$x_{i,new}^{t+1} = \begin{cases} x_i^{t+1} & |f(x_i^{t+1}) \leq f(x_i^t) \\ x_i^t & |otherwise \end{cases} \quad (5)$$

In this regard, $x_{i,new}^{t+1}$ newly produced butterfly for the next generation. and $f(x_i^{t+1})$ and $f(x_i^t)$ are the fitness level of butterflies x_i^t and x_i^{t+1} , respectively.

The structure of propellers in the M-dimensional space of features is shown in Fig. 3. To search in the butterfly-fire algorithm space, the features are coded as butterflies in Fig 3. A unique code is considered for each feature in all P of the

license. After defining the parameters, the fitness function will be calculated.

In order to escape from the local optimum and cover more search space of each graph, instead of selecting the worst vertex in each step, τ th worst vertex is selected; That is, the vertex whose rank k is calculated using eq. (3) is selected for replacement.

$$k = (1 + (n^{1-\tau} - 1)(rand)^{\frac{1}{1-\tau}}) \quad (6)$$

In this regard, k refers to the active number selected from the ordered list of ranks; n is the number of graph vertices and (rand) is the random number generation function in the interval [1, 0]. The value of parameter τ is fixed, and in this research, it is determined by trial-and-error method; For the algorithm limits the search space and, in other words, looks for the answer more strictly.

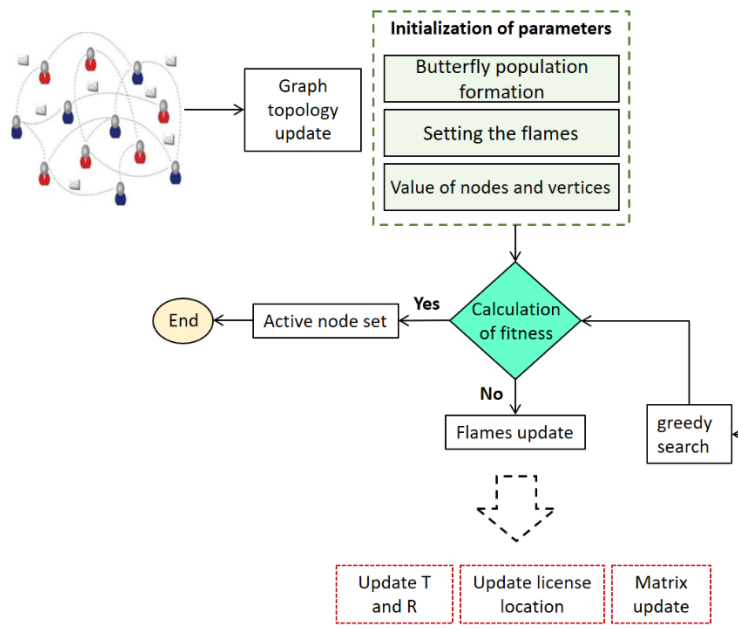


Fig. 1. Flowchart of the proposed algorithm.

```

Input: graph (V, C), Initializing the population
Output: Set of the Influence node
Update the number of flames (FlameNumber)
Initializing all the parameters
Initialize the population of moths
Calculate the objective values Equations (1)
for all moths
for all parameters
update r and t
Calculate D with respect to the corresponding moth by Equations (3)
Update the matrix M with respect to the corresponding moth by Equations (4) and (5).
Calculate the Influence of each node
end
calculate the objective values
Update j-th flames by greedy strategy as Equations (6)
Apply later flames
S the current best individual in the population.
end
Return S.
End
    
```

Fig. 2. Pseudo code of the proposed method.

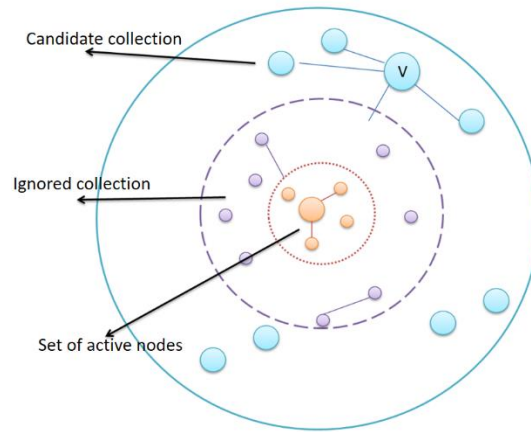


Fig. 3. The structure of nodes in the graph space.

IV. EVALUATION AND SIMULATION

Five community detection algorithms have been used to evaluate and compare the proposed method. The paper [19] presents a strategy called 2-Phase Community Detection (2PCD) to improve effective node detection algorithms in which a pre-processing step is added, and edges are weighted according to their centrality in the network topology. In the paper [4], an algorithm called Atom Stabilization Algorithm (ASA) has presented a new method for identifying influential nodes, which provides a new perspective for understanding the structure of nodes in complex networks. The algorithm presented in this paper is also used for evaluation. In this paper, the algorithms improved in research [4] and [19] have been used to evaluate the proposed method. To evaluate the results of the proposed method, the scale and quality criteria that are introduced in detail in the fourth section are used.

A. Measure Evaluation

In this section, for each of these data sets, the measure obtained by the proposed algorithm and other evaluated methods have been calculated. The relevant results are reported in Table II and Fig. 4.

From the analysis of Table II, it can be concluded that using the proposed algorithm has led to better results. The algorithm's performance has been tested on large datasets for which scaling optimization is increasingly difficult, and has shown satisfactory improvement. The proposed algorithm has improved by about 13.31% compared to the reference [19]. Compared to reference [4], it has improved significantly by 16.23.

TABLE II. THE OBTAINED MEASURE

No. of Data set	2PCD	ASA	CC-BF
1	0.7846	0.5795	0.8560
2	0.7031	0.6235	0.7349
3	0.6892	0.6768	0.8296
4	0.7303	0.6138	0.7612
5	0.7865	0.6895	0.8923
6	0.8032	0.6725	0.8514
7	0.7927	0.6614	0.8225
8	0.7823	0.6424	0.8368

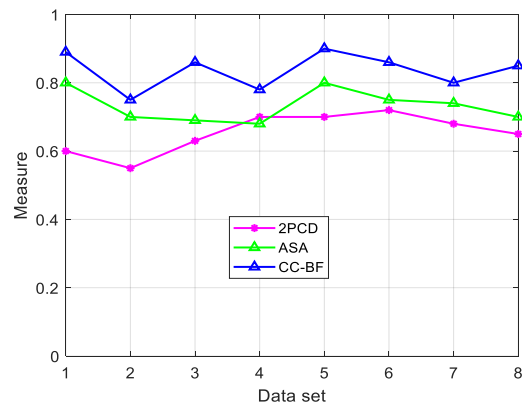


Fig. 4. Benchmark evaluation for 2PCD, ASA and CCBF methods.

According to the maximum scale obtained, the best values for parameters α_1 , α_2 and α_3 are obtained. These three parameters are used to control the relative importance of each feature. The obtained values are shown in Table III.

B. Quality Assessment

In this section, the quality of the communities created by the proposed strategy is analyzed by the NMI criterion. In this criterion, it is assumed that, according to graph G , an implicit truth is available in order to examine the communities. However, calculating NMI is challenging for real-life networks because no ground truth is usually available to assess what communities exist in G and their characteristics. Therefore, to perform experiments, a dataset should be considered in which the correct clustering of communities is specified. For this purpose, only the College Football Association of America dataset has been used in this section. The obtained results are shown in Table IV.

Quality value (NMI) ranges from 0 to 1, and higher values correspond to better algorithms. From the analysis of this table, we conclude that using the proposed algorithm has led to better results. The proposed algorithm has improved by approximately 3.7% compared to the algorithm presented in [4]. Compared to the algorithm [19], it has improved by about 4.7%. In the figure below, the computation time for 100 nodes is displayed.

TABLE III. THE BEST VALUES FOR PARAMETERS α_1 , α_2 AND α_3

Parameters	Value
α_1	0.64
α_2	0.81
α_3	0.44

TABLE IV. COMPARING THE QUALITY CRITERIA OBTAINED FOR THE PROPOSED ALGORITHM AND SIMILAR ALGORITHMS

Method	Value
2PCD	0.536508
ASA	0.679523
CC-BF	0.782158

C. Execution Time

In Fig. 5 and Table V, the results of the proposed method in this research are compared with similar methods. According to this table, it is clear that the proposed method has a real-time feature compared to similar methods.

Time and memory restrictions are increased in real-time applications. In the suggested method, a search is conducted for each active pattern to discover all active nodes with it before allowing any active users into the system. Each search's outcome is saved as a cluster. Upon receiving an alarm in real-time correlation, it is simple to locate earlier connected alerts by searching in the corresponding cluster. As a result, each node's processing time is cut down. The test results attest to the accuracy of the performance and the method's time-saving effectiveness. The method's benefit is that it executes in 5.3126 seconds on a home computer, proving it is real-time. Investigations have been conducted for the suggested method based on the source [23] for an online social network. The network of Twitter's "followership" is represented by this data. Network nodes represent users, and a link is created between two users when one of them is "followed". This network has 24256 nodes and 44135 edges. The average CC is 0.412, and the average network degree is 3.05. Table VI displays each network's kind and size.

In order to assess the consistency of CC-BF, each network in this study had CC-BF implemented 12 times, and the maximum, mean, and standard error of modularity values for those 12 implementations were determined. Fig. 6 displays the mean modularity values and standard errors.

TABLE V. COMPUTING TIME FOR 100 NODES

Method	Second(s)
2PCD	8.469144
ASA	7.562415
CC-BF	5.398828

For instance, CC-BF gives steady findings with 0% standard error for the highly sparse Facebook network, as shown in Fig. 6, and significantly beats all other methods. It is modular in terms of size. Cora performs poorly and generates a low modularity rating for this network. On the other hand, methods other than dolphin provide numbers for this network's modularity that are nearly identical. Jazz performs better than all other approaches for SCN, whereas CC-BF performs better than Ecoil, Protein, PGP, Cora, and Polbook.

Additionally, the Reality Mining dataset [24], which the Harvard Media Lab supplied, has been chosen to test the suggested methodology for the infection rate of users in social networks. 200 Harvard students' contact, proximity, location, and activity data from the academic years 2006–2008 are included in the reality mining dataset.

In each test on this dataset, 0.03% of users are infected with a worm and worm propagation is simulated over three days. The CC-BF method's infection rate in each test is contrasted with the social-based 2PCD and ASA approaches. The percentage of infected users over non-infected users is used to compute the infection rate. Similar techniques yield 200 (for a static network), 300 (for a dynamic network), and 350 (for a network) clusters, respectively. The warning threshold β is established at 3, 15, and 25%, respectively, for each value of m .

For stability, each experiment is run 2,000 times. The experiment's findings for three distinct values of m and β are displayed in Fig. 7, 8, and 9. First, it is evident that in order to meet the anticipated infection rate, more people must be patched the longer one waits (, the higher the warning threshold). For instance, with $m = 200$ clusters and an anticipated contamination rate of 0.4, we should allocate patches to more than 20% of the users when $\beta = 12\%$ and to users who would be sent when $\beta = 3\%$. When $\beta = 25\%$, approximately 94% of all users have influence.

The second finding demonstrates that, in the static version of the social network depicted in Fig. 7, the proposed method outperforms the social-based methods 2PCD and ASA in terms of contamination rate. For instance, the proposed method's contamination rate is 7% to 12% lower than existing algorithms with comparable goals. The number of clusters m changes when new users join the network and form new social connections, and the infection rate of the social-based technique is updated using the cluster sizes of 300 and 350 as well as the warning threshold. $\beta = 3\%$, 15% and 25%, respectively. Fig. 8 and 9 illustrate how the suggested method, which has the ability to quickly and adaptively update the network community structure, achieves a higher infection rate than competing methods while also having much lower computational costs and execution times.

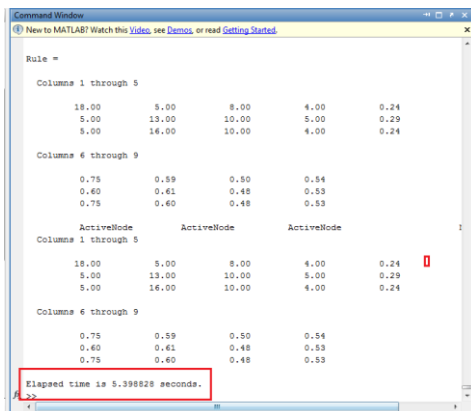


Fig. 5. The computational time for 100 nodes.

TABLE VI. EMPIRICAL AND PRACTICAL DATE SETS ARE UTILIZED TO ASSESS CC-BF

Data set	Type	Edge	Nodes	Avg. ccpppp	Avg. degrees	Avg. p.length
BA-1000	Social	74.1	32.3	0.5586	4.3586	2.2876
FF-256	Online-Social	151.05	58.9	0.28785	4.87255	3.18915
LFR256	Collaboration	582.35	109.25	0.38285	10.12795	2.3826
GR512	Collaboration	418.95	99.75	0.4636	7.98	2.92505
GR1000	Social	5209.8	188.1	0.60135	26.31215	2.12325
BA-512	Social	493.05	397.1	0.19855	2.35885	4.58185
E.coli	Collaboration	2604.9	1509.55	0.8341	3.27845	5.53185
WS-512	Synthetic	2831.95	2743.6	0.76285	1.9608	3.67365
Dolphins	Synthetic	8310.6	3537.8	0.2071	4.4631	4.99605
Football	Online-Social	13771.2	4979.9	0.65265	5.25445	5.74655
BA-256	Online-Social	23100.2	10146	0.418	4.275	7.11075
WS-256	Biological	31189.45	22201.5	0.3154	2.679	0.59945
Facebook	Synthetic	1945.6	243.2	0.3287	15.2	2.5764
Protein	Social	1050.7	243.2	0.51775	6.76115	3.477
FF-512	Online-Social	1983.6	486.4	0.5434	7.2656	3.6784
FF-1000	Synthetic	3802.85	950	0.53485	7.6722	3.9064
Twitter	Synthetic	242.25	243.2	0.95	1.8924	5.47295
WS-1000	Collaboration	485.45	486.4	0.95	1.8962	6.1123
NetScience	Synthetic	949.05	950	0.95	1.8981	7.08795
Scientific Co.	Biological	780.9	243.2	0.54815	6.1009	3.53305
PGP	Biological	3156.85	486.4	0.57665	12.331	3.2566
Jazz	Biological	10719.8	950	0.56905	21.4396	2.98015
GR256	Social	1216	243.2	0.48355	9.5	3.28415
Karate	Collaboration	2432	486.4	0.46835	9.5	3.76675
Polbooks	Online-Social	4750	950	0.47405	9.5	4.2617

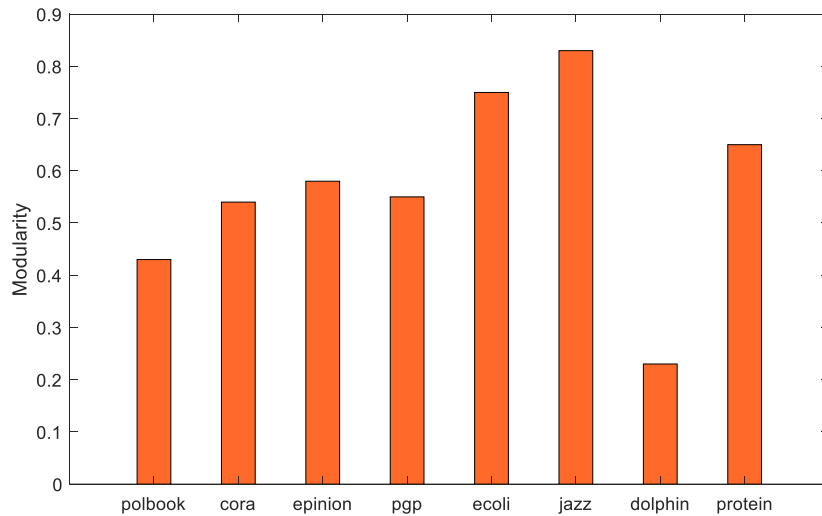
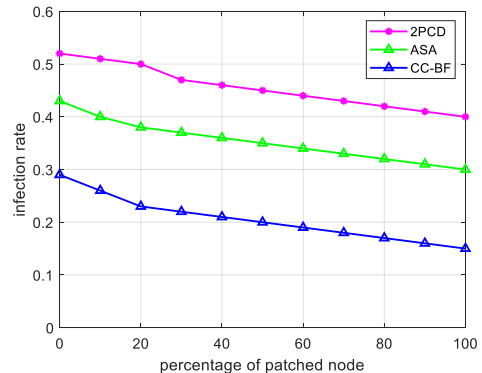
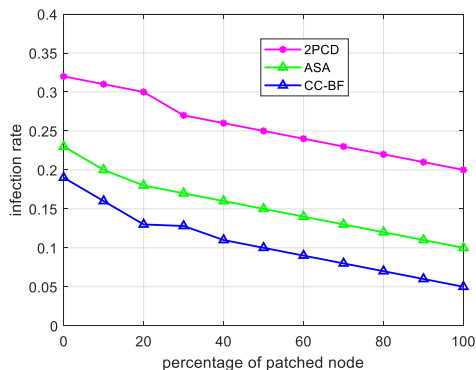


Fig. 6. Standard errors and modular mean values for 12 CC-BF runs across several networks.



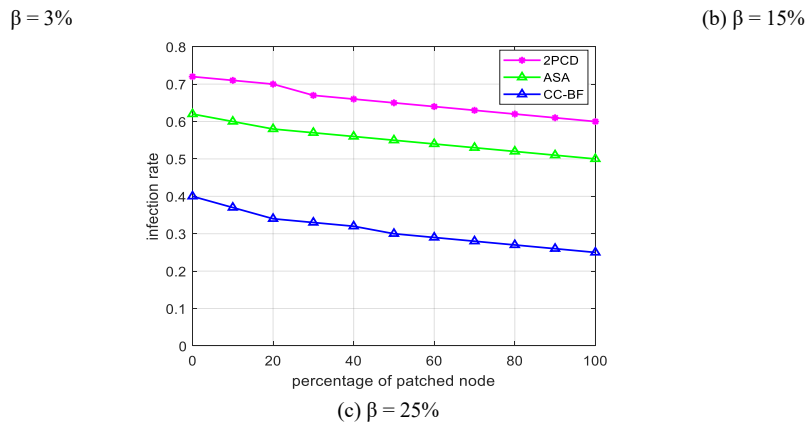


Fig. 7. A static network with $k = 200$ clusters and infection rates.

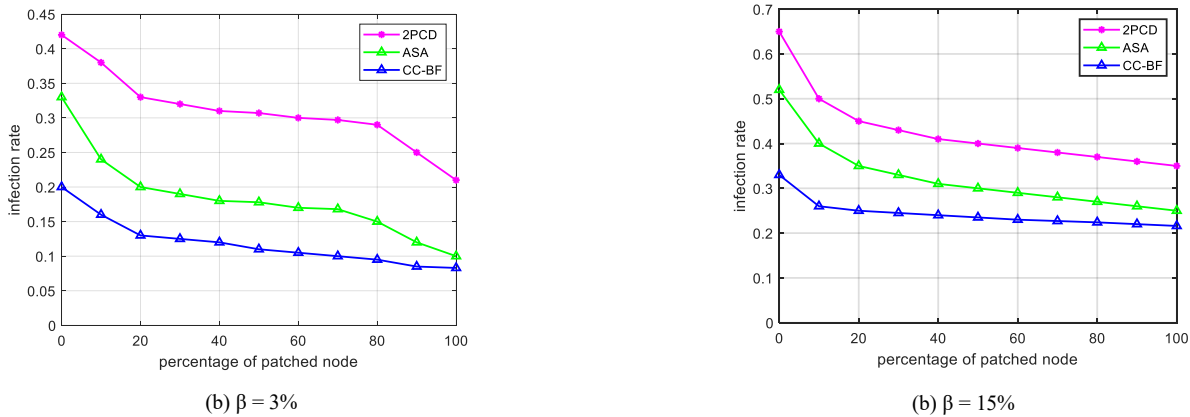


Fig. 8. A static network with $k = 300$ clusters and infection rates.

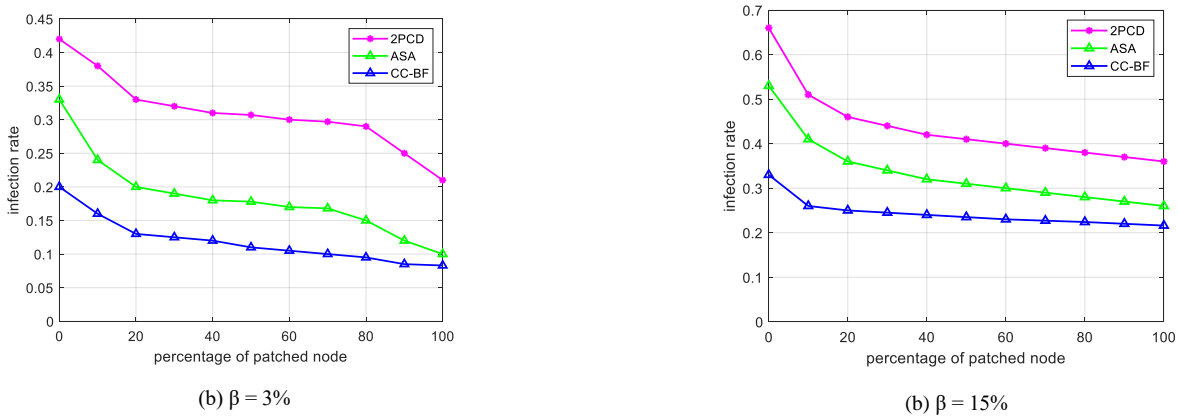


Fig. 9. A static network with $k = 350$ clusters and infection rates.

V. CONCLUSION

Investigating community recognition in social networks is an important issue in many fields and disciplines, such as marketing and information technology. Community detection in social networks can be considered a graph clustering problem, where each set corresponds to a cluster in the graph. The goal of conventional community detection methods is to partition a graph so that each node belongs to exactly one cluster. A community can be defined as a group of individuals close to each other compared to other entities in the dataset.

The proposed algorithm is checked and compared with the evaluated algorithms. This comparison is done according to the introduced criteria. To provide more accurate and complete results, five data sets have been used. Examining the results shows that the proposed method has led to better results. The performance of the algorithm has been tested on large datasets, for which scaling optimization is increasingly difficult and has shown satisfactory improvement. Among the advantages of NMI compared to other related methods, we can mention the competition with the latest technologies in the accuracy and

scale of data, the ability to execute and implement the system in the real world, and the ability to perform all the steps of the proposed system online. The results show that the proposed algorithm has improved by 23.6% compared to previous similar work. In order to improve the quality of detecting effective nodes in future works, it is possible to mention the discovery of the semantic relationship between nodes and the virtual networks between them. Hidden connections between nodes are introduced as virtual associations. Discovering virtual associations between nodes can help algorithms based on participation in the detection process. Also, tagging the content of messages sent between nodes can help to discover nodes.

REFERENCES

- [1] A. Biswas, S. Khandelwal, and B. Biswas, "Community detection in networks using atom stabilization algorithm," 2017: IEEE, pp. 89-93.
- [2] P. Chunaev, "Community detection in node-attributed social networks: a survey," *Computer Science Review*, vol. 37, p. 100286, 2020.
- [3] M. Mazza, G. Cola, and M. Tesconi, "Modularity-based approach for tracking communities in dynamic social networks," *arXiv preprint arXiv:2302.12759*, 2023.
- [4] N. Dakiche, F. B.-S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," *Information Processing & Management*, vol. 56, no. 3, pp. 1084-1102, 2019.
- [5] R. Djerbi, R. Imache, and M. Amad, "Communities' Detection in Social Networks: State of the art and perspectives," 2018: IEEE, pp. 1-6.
- [6] M. Contisciani, H. Safdari, and C. De Bacco, "Community detection and reciprocity in networks by jointly modelling pairs of edges," *Journal of Complex Networks*, vol. 10, no. 4, p. cnac034, 2022.
- [7] Trik, M., Pour Mozafari, S., & Bidgoli, A. M. (2021). An adaptive routing strategy to reduce energy consumption in network on chip. *Journal of Advances in Computer Research*, 12(3), 13-26.
- [8] Yanwei Zhao, Ben Niu, Guangdeng Zong, Xudong Zhao, Khalid H. Alharbi. Neural network-based adaptive optimal containment control for non-affine nonlinear multi-agent systems within an identifier-actor-critic framework, *Journal of the Franklin Institute*. 360 (12), pp.8118-8143, 2023.
- [9] S. Guo, X. Zhao, H. Wang, N. Xu, Distributed consensus of heterogeneous switched nonlinear multiagent systems with input quantization and dos attacks, *Applied Mathematics and Computation* 456 (2023) 128127.
- [10] Khezri, E., Zeinali, E., & Sargolzaey, H. (2023). SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks. *Plos one*, 18(4), e0282031.
- [11] Arefanjazi, H., Ataei, M., Ekramian, M., & Montazeri, A. (2023). A Robust Distributed Observer Design for Lipschitz Nonlinear Systems With Time-Varying Switching Topology. *Journal of the Franklin Institute*.
- [12] A. A. Paracha, J. Arshad, and M. M. Khan, "SUS You're SUS!—Identifying influencer hackers on dark web social networks," *Computers and Electrical Engineering*, vol. 107, p. 108627, 2023.
- [13] Fabin Cheng, Ben Niu, Ning Xu, Xudong Zhao, and Adil M. Ahmad. Fault Detection and Performance Recovery Design With Deferred Actuator Replacement Via A Low-Computation Method, *IEEE Transactions on Automation Science and Engineering*, DOI: 10.1109/TASE.2023.3300723, 2023.
- [14] Khezri, E., Zeinali, E., & Sargolzaey, H. (2022). A novel highway routing protocol in vehicular ad hoc networks using VMaSC-LTE and DBA-MAC protocols. *Wireless Communications and Mobile Computing*, 2022.
- [15] Chen Cao, Jianhua Wang, Devin Kwok, Zilong Zhang, Feifei Cui, Da Zhao, Mulin Jun Li, Quan Zou. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. *Nucleic Acids Research*.2022, 50(D1): D1123-D1130.
- [16] FanghuaTang, Huanqing Wang, Liang Zhang, Ning Xu, Adil M.Ahmad. Adaptive optimized consensus control for a class of nonlinear multi-agent systems with asymmetric input saturation constraints and hybrid faults. *Communications in Nonlinear Science and Numerical Simulation*, 126: 107446, 2023.
- [17] Sai Huang; Guangdeng Zong; Huanqing Wang; Xudong Zhao; K. H. Alharbi. Command Filter-Based Adaptive Fuzzy Self-Triggered Control for MIMO Nonlinear Systems With Time-Varying Full-State Constraints. *International Journal of Fuzzy Systems*, 2023, <https://doi.org/10.1007/s40815-023-01560-8>.
- [18] E. Jokar, M. Mosleh, and M. Kheyrandish, "Overlapping community detection in complex networks using fuzzy theory, balanced link density, and label propagation," *Expert Systems*, vol. 39, no. 5, p. e12921, 2022.
- [19] Khezri, E., & Zeinali, E. (2021). A review on highway routing protocols in vehicular ad hoc networks. *SN Computer Science*, 2, 1-22.
- [20] S. Tschitschek, A. Singla, M. Gomez Rodriguez, A. Merchant, and A. Krause, "Fake news detection in social networks via crowd signals," 2018, pp. 517-524.
- [21] M. Trik, H. Akhavan, A.M. Bidgoli, A.M.N.G. Molk, H. Vashani, S.P. Mozaffari, A new adaptive selection strategy for reducing latency in networks on chip, *Integration*. 89 (2023) 9–24.
- [22] Khalafi, M., & Boob, D. (2023, July). Accelerated Primal-Dual Methods for Convex-Strongly-Concave Saddle Point Problems. In *International Conference on Machine Learning* (pp. 16250-16270). PMLR.
- [23] C.-C. Ni, Y.-Y. Lin, F. Luo, and J. Gao, "Community detection on networks with Ricci flow," *Scientific reports*, vol. 9, no. 1, pp. 1-12, 2019.
- [24] Behzad, A., Wakkary, R., Oogjes, D., Zhong, C., & Lin, H. (2022, April). Iterating through Feeling-with Nonhuman Things: Exploring repertoires for design iteration in more-than-human design. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (pp. 1-6).
- [25] M. M. D. Khomami, A. Rezvanian, M. R. Meybodi, and A. Bagheri, "CFIN: A community-based algorithm for finding influential nodes in complex social networks," *The Journal of Supercomputing*, vol. 77, pp. 2207-2236, 2021.
- [26] M. Samiei, A. Hassani, S. Sarspy, I.E. Komari, M. Trik, F. Hassanpour, Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare, *J Cancer Res Clin Oncol*. (2023) 1–15.
- [27] H. Li, Q. Shang, and Y. Deng, "A generalized gravity model for influential spreaders identification in complex networks," *Chaos, Solitons & Fractals*, vol. 143, p. 110456, 2021.
- [28] Wakkary, R., Oogjes, D., & Behzad, A. (2022). Two Years or More of Co-speculation: Polylogues of Philosophers, Designers, and a Tilting Bowl. *ACM Transactions on Computer-Human Interaction*, 29(5), 1-44.
- [29] M. Trik, A.M.N.G. Molk, F. Ghasemi, P. Pouryeganeh, A hybrid selection strategy based on traffic analysis for improving performance in networks on chip, *J Sens*. 2022 (2022)
- [30] M. Al Assad and N. Agarwal, "Contextualizing focal structure analysis in social networks," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 103, 2022.
- [31] un, J., Zhang, Y., & Trik, M. (2022). PBPHS: a profile-based predictive handover strategy for 5G networks. *Cybernetics and Systems*, 1-22.
- [32] A. R. Costa and C. G. Ralha, "AC2CD: An actor–critic architecture for community detection in dynamic social networks," *Knowledge-Based Systems*, vol. 261, p. 110202, 2023.
- [33] M. R. HabibAgahi, M. A. M. A. Kermani, and M. Maghsoudi, "On the Co-authorship network analysis in the Process Mining research Community: A social network analysis perspective," *Expert Systems with Applications*, vol. 206, p. 117853, 2022.
- [34] M. Al Assad and N. Agarwal, "A Systematic Approach for Contextualizing Focal Structure Analysis in Social Networks," 2022: Springer, pp. 46-56.

Motor Imagery EEG Signals Marginal Time Coherence Analysis for Brain-Computer Interface

Md. Sujan Ali, Mst. Jannatul Ferdous

Department of Computer Science and Engineering,
Jatiya Kabi Kazi Nazrul Islam University, Trishal, Mymensingh-2224, Bangladesh

Abstract—The synchronization of neural activity in the human brain has great significance for coordinating its various cognitive functions. It changes throughout time and in response to frequency. The activity is measured in terms of brain signals, like an electroencephalogram (EEG). The time-frequency (TF) synchronization among several EEG channels is measured in this research using an efficient approach. Most frequently, the windowed Fourier transforms-short-time Fourier transform (STFT), as well as wavelet transform (WT), and are used to measure the TF coherence. The information provided by these model-based methods in the TF domain is insufficient. The proposed synchro squeezing transform (SST)-based TF representation is a data-adaptive approach for resolving the problem of the traditional one. It enables more perfect estimation and better tracking of TF components. The SST generates a clearly defined TF depiction because of its data flexibility and frequency reassessment capabilities. Furthermore, a non-identical smoothing operator is used to smooth the TF coherence, which enhances the statistical consistency of neural synchronization. The experiment is run using both simulated and actual EEG data. The outcomes show that the suggested SST-dependent system performs significantly better than the previously mentioned traditional approaches. As a result, the coherences dependent on the suggested approach clearly distinguish between various forms of motor imagery movement. The TF coherence can be used to measure the interdependencies of neural activities.

Keywords—Brain-Computer Interface (BCI); Electroencephalogram (EEG); Short-time Fourier Transform (STFT); Synchrosqueezing Transform (SST); time-frequency coherence

I. INTRODUCTION

Brain signals (EEG) can be used to build the Brain Computer Interface (BCI), which is a more convenient and affordable method. EEG signals are captured by spatially scattered scalp sensors. The connections between the various areas of the brain, which is the primary organ of the nervous system, are becoming important in BCI research. The various sensors record the EEG signals, and coherence analysis is used to determine how coherent the signals are [1, 2, 3]. Coherence is typically estimated using spectral methods such as a Fourier or wavelet [4] transform. Coherence analysis is challenging to implement because cerebral activity signals are inherently non-stationary. Although the time-dependent Fourier transform (STFT) is one approach to solving the issue, it has not been completely effective for the aforementioned reasons: one cannot guarantee the stationarity of brain signals during each brief time interval and two the Heisenberg uncertainty

principle limits the resolution of time-frequency representation. Despite being a data-adaptive signal analysis technique, the mother wavelet basis function is used in the wavelet transform to decompose signals. The approach also has difficulty with time-frequency resolution, where the resolution of the frequency is greater at low frequencies and less so at high frequencies. Also, this method is founded on choosing a mother wavelet. Because the mother wavelet was arbitrarily chosen without being matched to the analyzing signal, that led to an inaccurate and irreversible breakdown.

When combined with the continuous wavelet transform (CWT), the technique, known as synchrosqueezing transform (SST) [5], produced astoundingly precise time-frequency depictions of nonstationary as well as nonlinear data. This aspect of SST addresses the drawbacks of linear perception time-frequency techniques, such as windowed Fourier transforms (STFT) as well as continuous wavelet transforms. The synchrosqueezing transformation focuses the coefficient values around the frequency response graph of the tuned oscillations by dispersing the STFT and CWT strengths [6]. The frequency redistribution approach used in time-frequency representation [7] improves the proper location of instantaneous amplitude in the time and frequency domains.

Since neural synchronization is characterized by several frequency bands but is anticipated to change over time, TF coherence is typically employed for measuring it. Smoothing the cross as well as auto-spectra between the signals is essential since noise has a significant impact on coherence. One of the following techniques is used to execute the smoothing operation: periodogram smoothing can be accomplished in one of three ways: (i) Periodogram smoothing through ensemble averaging using the WOSA (Welch's overlapped segment averaging) technique; (ii) the temporal or frequency domains may be smoothed separately or together [8, 9]; and (iii) By averaging a collection of spectra generated using various orthogonal taper functions, cross and auto spectra are smoothed. The cross and auto spectra are typically smoothed with the same smoothing agents in all of the approaches mentioned above to estimate TF consistency. The employment of the same smoothing operator constrained the coherence to the range [0, 1] since the TF coherence satisfies the Cauchy-Schwarz inequality. Furthermore, the estimator using the same operation fails when the smoothing coefficient rises to one. Selected auto spectra smoothing can be used to get the improved temporal resolution. However, since the cross spectra are therefore not flattened when non-identical smoothing operators are used, the bias of the

estimator cannot reach one [8]. As a result, the estimator has better time resolution. In order to properly depict TF consistency and uncover weak correlations among signals, non-identical smoothing agents may be used. Bispectrum-based channel selection (BCS) was employed in this study [10] for MI-based BCIs. In this paper [11], the performance of the BCI model may be considerably impacted by using different time segments for training the data. They recommend against using any other temporal data as training data besides that utilized for motor imaging. For BCI Competition IV 2a and 2b, models using machine learning and deep learning suggested a potential improvement in visual display time, categorization efficiency. It was argued that models might be picking up more visual information. In fact, during the visual presentation, spatial topography revealed activation of the visual cortex. In the research [12], MI classification using EEG signals is accomplished using a supervised feature selection method. Another work [25] suggests a technique for producing a spatio spectral feature representation that can maintain the multivariate information of EEG data. In particular, subject-optimized and subject-independent spectrum filters were combined, and the filtered data were then stacked into tensors to create 3-D feature maps. In order to automatically choose the best frequency bands based on MIF [13], the MIFCSP method combines multivariate iterative filtering (MIF) and CSP. This method may then be used to extract discriminant features.

The SST approach is employed in the present study to calculate the TF coherence of brain signals, as well as the coherence is therefore subjected to a non-identical smoothing procedure. Moreover, the same analysis is carried out using the short-time Fourier transform rather than the SST. With synthetic and actual EEG signals, both findings are validated. The SST-dependent TF consistency outperforms the STFT-dependent technique, according to the observation in both synthetic and real data. The following is how the paper is set up: The time-frequency representation techniques, such as STFT and SST, are covered in Section II, along with the consistency in the TF domain in Section II, the synchronized transformation research findings in Section III, and discussion and some closing thoughts in Sections IV and V, respectively.

II. METHODS

The neuronal synchronization changes both over time and with frequency. Any signal's energy is described as a function of both times as well as frequency by the time-frequency representation (TFR). It converts a single-dimensional time-series signal through a double-dimensional function integrating frequency and period. The TFR space value gives a sense of which spectral components are present. Non-stationary or time-varying signals can be analyzed and created using the TFR.

A. Windowed Fourier Transform

An approach that works well for the TF characterization of non-stationary EEG signals is the short-time Fourier transform (STFT). A kind of trade-off between a signal's time and frequency is made by the STFT, which contains all the data on

frequency variations with the period. A signal event's timing and frequency are also disclosed through this information. The signal is broken up into manageable chunks for the duration of the STFT, and it may be expected that each of these chunks will remain stationary. A window function (w) is selected in order to achieve this. This window must have the same width as the area of the signal when the normality of the data is guaranteed. The definition of the STFT for a non-stationary signal $s(t)$ is

$$\Psi(t, f) = \int_{-\infty}^{\infty} [s(t) \cdot w^*(t-t')] \cdot e^{-2\pi f t} dt \quad (1)$$

where even the window function $w(t)$ and the complex conjugate $*$ are both present. The signal's STFT is the signal's Fourier transform times a tapering function [27].

In the interests of demonstrating the time-frequency depiction, a noise-free artificial signal is created that is called $\delta(t)$ by chaining three sinusoids $s1(t)$, $s2(t)$, and $s3(t)$ with frequencies of 10Hz, 5Hz, and 20Hz, respectively, with the formula as $\delta(t) = [s1(t) s2(t) s3(t)]$. For sampling, 500 hertz (Hz) is employed. The STFT-based synthetic signal and TFR are shown in Fig. 1(a) and 1(b), respectively. A Hamming window with a 50% overlap and a length of 256 is employed in the STFT. Although with poor resolution, the STFT can distinguish between the three factors.

B. Synchronized Transformation

A useful method for the Continuous Wavelet Transform (CWT) is the Synchrosqueezing Transform (SST). This method is employed to concentrate the frequency elements of non-stationary signals in the TF space. The CWT successfully creates a high-resolution TF representation. In Fig.1, the SST-dependent TF visualization derived from the artificial signal $\delta(t)$ is displayed. The right CWT scales are used for discretization, and a bump mother wavelet is used to achieve SST. It has been highlighted that STFT-based TF space suffers from extremely poor frequency resolution and reduced temporal resolution owing to the employment of the window function. Using a set of wavelets, which are time-frequency filters, the CWT method detects oscillatory elements in a signal. The CWT is used to create wavelets from a successive time function. In the following form, a signal $s(t)$ is convolved with a mother wavelet $\Phi(t)$, which is a finite oscillatory consequence.

$$Z(p, q) = \frac{1}{|p|^{1/2}} \int_{-\infty}^{\infty} \Phi\left(\frac{t-q}{p}\right) s(t) dt \quad (2)$$

If each scale-time duet's (p, q) wavelet coefficients are represented by $Z(p, q)$, then it is possible to determine the instantaneous frequency $\omega_s(p, q)$ by using the formula

$$\omega_s(p, q) = -iZ(p, q)^{-1} \frac{\partial Z(p, q)}{\partial q} \quad (3)$$

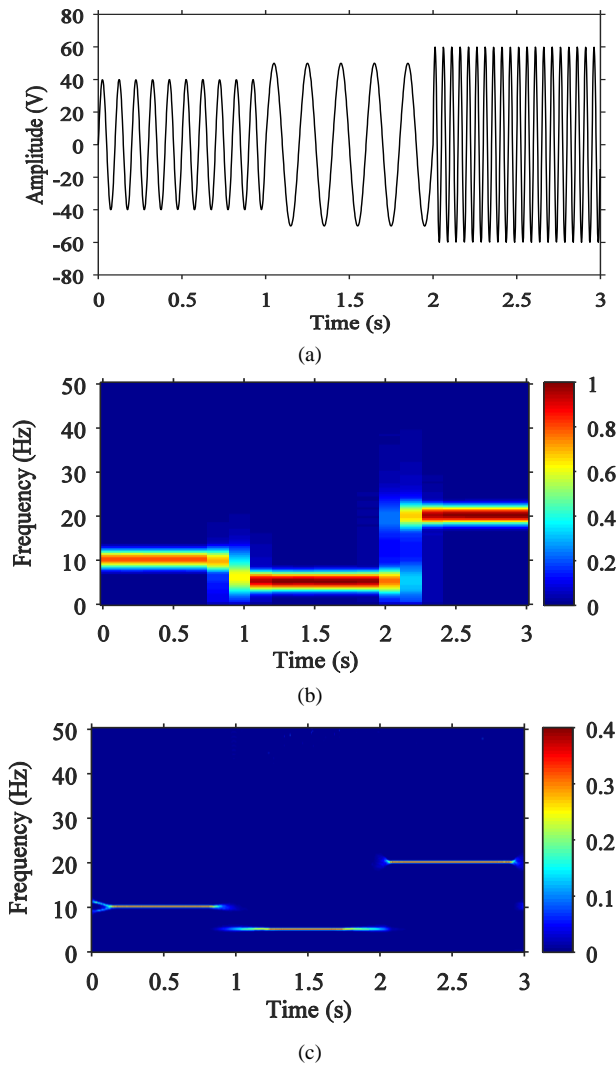


Fig. 1. A synthetic signal $\delta(t)$ with three sinusoids is represented by TF utilizing (a) the simulated signal $\delta(t)$ and (b) STFT as well as (c) SST.

With TF representation, the information from the time-scale frame is translated to the time-frequency frame. During the synchrosqueezing procedure, each value is changed to $(q, \omega_s(p, q))$ [5]. It is able to have a scaling step because p and q are distinct numbers; for each, p_k , where $\omega_s(p, q)$ is calculated. When projecting from the time-scale frame to the time-frequency frame $(q, p) \rightarrow (q, \omega_{inst}(p, q))$, the SST $\Gamma(\omega_l, q)$ is only calculated [11] in the centres ω_l located in the spectral region $[\omega_l - \Delta\omega/2, \omega_l + \Delta\omega/2]$ with $\Delta\omega = \omega_l - \omega_{l-1}$.

$$\Gamma(\omega_l, q) = \sum_{p_k: |\omega_s(p_k, q) - \omega_l| \leq \Delta\omega/2} Z(p_k, q) p^{-3/2} \Delta p_k \quad (4)$$

Eq. (4) demonstrates that only the frequency (or scale) axis is synchrosqueezed in the TF representation of the signal [14, 28]. To obtain a focused image over the time-frequency plane for the SST, the CWT coefficients are reallocated [15]. The instantaneous frequencies are then taken from this image.

C. Coherence Evaluation

Effective communication between two parties can be achieved through coherence. Cohesiveness in neuroscience describes the systematic constancy among two neuronal cells. The establishment of more or less uniformity between oscillating modulations in different neurons' brain activity is known as neuronal coordination. Synchronization has a substantial impact on how the different neuronal regions synchronize their stimulatory behavior [16, 17, 18].

D. Frequency Coherence

A common technique for assessing consistency within brain waves is frequency consistency. Frequency coherence's major benefits include being highly implicit, hard, and noise-resistant while allowing for a quick overview of pertinent consistent frequencies in the sample [19]. The frequency coherence is a measure of how well multiple signals' cross-spectral levels hold up when normalized with respective auto-spectral levels. As functions of frequency, consider x and y , two stationary random processes. According to [26], the familiar consistency function of x , as well as y , is as follows:

$$|C_{x,y}(f)| = \frac{|J_{x,y}(f)|}{\sqrt{J_{x,x}(f)J_{y,y}(f)}} \quad (5)$$

wherever $J_{x,y}(f)$ is the cross-spectral density among the two processes. $J_{x,x}(f)$ and $J_{y,y}(f)$ are the auto spectral density functions of x as well as y , respectively, at frequency, f . The EEG is a non-stationary signal; hence the conventional coherence function is insufficient.

E. Coherence of Time and Frequency

Typically, coherence analysis only works with stationary signals since it calculates the relationship between two signals throughout the frequency region. Consequently, much like with non-stationary signals, traditional coherence analysis is unable to reveal the temporal features of EEG [20]. The sequential relationship among both processes in the time-frequency dimension is measured using an advanced ruling technique. The TF consistency has been employed to gauge the synchronization of cortical activity in the brain-computer interaction motor imagery experiment. The coherence characteristic of the TF is described as

$$|C_{x,y}(t, f)| = \frac{|J_{x,y}(t, f)|}{\sqrt{J_{x,x}(t, f)J_{y,y}(t, f)}} \quad (6)$$

At this point, $t = 1, 2, \dots, T$; $f = 1, 2, \dots$, is the distinct frequency and the signal is divided into T segments. The measurements of the sectional as well as auto-spectral concentrations are

$$J_{x,y}(t, f) = X(t, f)Y^*(t, f)$$

$$J_{x,x}(t, f) = |X^2(t, f)|, \quad J_{y,y}(t, f) = |Y^2(t, f)| \quad (7)$$

where $X(t, f)$ and $Y(t, f)$ are the respective x also y of TF transforms coefficients, besides $Y^*(t, f)$ is the complex quantity of $Y(t, f)$.

The TF consistency definitions are simple and also use a method akin to the Fourier analysis. Based on the spectrogram approach, which involves averaging the signal segments to arrive at the estimates, the spectra and the frequency coherence in the Fourier analysis can be calculated. As both time and frequency have two dimensions, the time-frequency consistency encounters challenges throughout averaging. SST, which performs better than STFT, is used throughout this study to compute the TF translation parameters accompanied by TF consistency

F. Smoothing Impacts on TF Coherence

To get rid of noise, utilize the smoothing operator, and a convolution operator. The operators for smoothing cross- and auto-spectral densities can be the same or different. The employment of non-identical operators, as opposed to identical operators, produces time-frequency consistency that is unrestricted to $[0, 1]$ and, as a result, improves temporal resolution [8]. Smoothing both time and frequency is necessary to increase the TF coherence’s constancy. Averaging a number of orthogonal-based spectrum estimations, such as those obtained using multi-taper methods, can serve as the smoothing operator. They are typically employed for amplitude- and auto-spectral densities. The non-identical smoothing agents are two-dimensional in both time and frequency [21] or a single-dimensional function of time. Thus, the standard magnitude squared TF coherence is calculated as [8]

$$|C_{x,y}(t, f)|^2 = \frac{|J_{x,y}(t, f) \otimes w[\phi]|^2}{\{J_{x,x}(t, f) \otimes w[\phi]\} \{J_{y,y}(t, f) \otimes w[\phi]\}} \quad (8)$$

Here, $w[\phi]$ also $w[\varphi]$ remain two different ($w[\phi] \neq w[\varphi]$) leveling windows of cross-spectral concentration and auto

spectral concentration, accordingly, and \otimes denote the convolution operator. The impacts of smoothing in TF consistency are demonstrated in Fig. 2, where two artificial signals $x_1 = [\sin(2\pi f_1 t) \sin(2\pi f_2 t)]$, $x_2 = [\sin(2\pi f_1 t) \sin(2\pi f_2 t)]$ with $f_1 = 5\text{Hz}$ also $f_2 = 10\text{Hz}$ and their TF consistency remain accessible. According to Fig. 2(a) and 2(d), respectively, the individual sinusoids that make up x_1 and x_2 have various temporal lengths.

As shown in Fig. 2(e), and 2(f), smoothing operations are shown to increase the TF coherence’s representativeness and clarity for both STFT and SST-based methods. On the other hand, as seen in Fig. 2(b) and 2(c), when the smoothing procedure is not carried out, a significant amount of irrelevant coherence is introduced. Hence, the measurement of time-frequency coherence is enhanced by utilizing diverse smoothing agents. Several 2-D Gaussian smoothing windows with various lengths are employed in this research. The kernel’s height in hertz and width in seconds are denoted by h and d , respectively, to reflect the window length $w = [h \ d]$.

G. Proposed Algorithm for TF Coherence

The steps that make up the suggested technique for calculating the time-frequency coherence amongst two signals dependent on SST are as follows:

- 1) Choose two EEG channels or two brain signals at random.
- 2) The time-frequency coefficients can be obtained by applying the SST to each individual signal.
- 3) The cross and auto spectral distributions should be calculated using the SST coefficients.
- 4) Using two acceptable non-identical (various window lengths) smoothing processes, amplify the cross and auto spectral densities.
- 5) Lastly, use Eq. (8) to get the time-frequency coherence by using smoothed auto and cross-spectral densities.

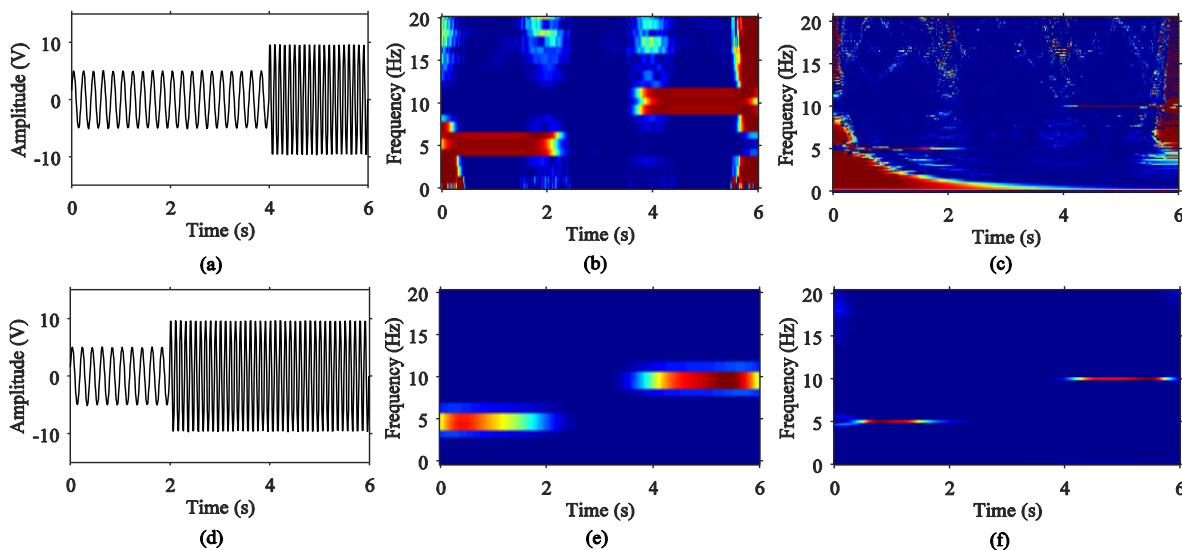


Fig. 2. The impact of smoothing operations on the TF consistency among the artificial signals x_1 (a) as well as x_2 (d). (b) STFT and (c) SST dependent TF lacking levelling, STFT and I STFT and (f) SST-dependent TF consistency through smoothing.

III. RESULTS

Both real EEG data and synthetic signals are utilized to evaluate the performance of the proposed SST-based time-frequency consistency. The outcomes are contrasted with time-frequency consistency dependent on STFT. With regard to STFT, a hamming window of length 100 is employed. Then, using non-identical smoothing windows, the spectral coefficients are smoothed for the calculation of TF coherence. The cross- and auto-spectral concentrations are levelled with Gaussian smoothing windows with lengths of w [2 1] and w [10 1], respectively. To execute the SST, a mother wavelet with bumps is utilized, and then the CWT scales' discretization is set at 32. The cross-spectral density is smoothed over in the SST because the Gaussian smoothing windows are w [3 1] and w [50 1] long. The Gaussian smoothing windows in the SST have lengths of w [3 1] and w [50 1], meaning that the cross and auto spectral densities are flattened across TF areas of 3 Hz and 1 s, respectively.

1) *Synthetic data:* Three sinusoids with frequencies of 5 Hz, 6 Hz, and 10 Hz are added together with a sampling frequency of 100 Hz to produce the trio of non-generated

signals X, Y, and Z. As shown in Fig. 3, each of the simulated signals are made up of such three signals with various temporal alignments. The distinct synthetic waveforms X, Y, and Z are then each polluted with 5 dB, 0 dB, and -5 dB of Gaussian noise, accordingly. Fig. 4 and 5 show the time-frequency consistencies within each couple of artificial signals produced by STFT and SST, separately. The consistency among the signals Y and Z (5Hz and 6Hz frequency) is displayed in Fig. 4. while Fig. 5 shows the cohesiveness between the exact same pair of signals is separated in a more pronounced manner, they overlapped each other. Fig. 6, which shows the marginal frequency coherences of two approaches, exemplifies the phenomena clearly (STFT and SST). If $f=1,2,\dots,F$, the definition of the marginal frequency coherence is $\tilde{C}_{x,y}(f) = \sum_{t=1}^T |C_{x,y}(t,f)|^2$. In STFT, values for closer frequency coherence values overlap, whereas, with SST, the coherence of each individual frequency component is strongly represented. It is believed that the SST-based technique has better resolution than the STFT-based time-frequency coherence technique.

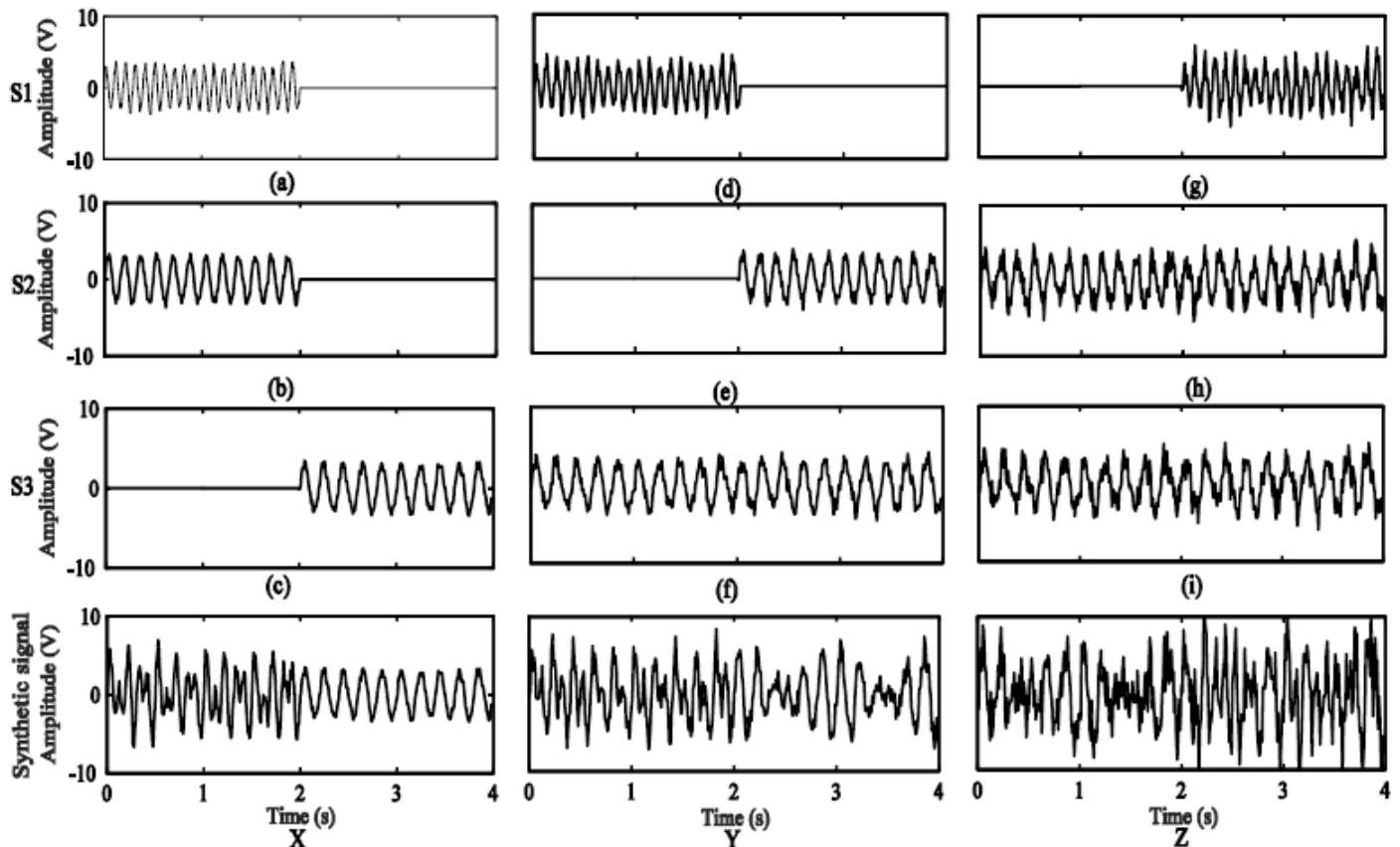


Fig. 3. Development of a multiple non-stationary [X, Y, Z] signal. Three separate frequencies sinusoids are present within the initial three rows (S1, S2, as well as S3). The three sinusoids in S1, S2, and S3 have, correspondingly, 10 Hz, 6 Hz, and 5 Hz frequency properties. To produce the artificial signals X, Y, and Z, alternative time alignments of the sinusoids are used. 5dB, 0dB and -5dB noises are inserted to sinusoids (a) to (c), (d) to (f) and (g) to (i) correspondingly. The synthesized signal in the fourth row is made up of the three sinusoids; $X=(a)+(b)+(c)$, $Y=(d)+(f)$ and $Z=(g)+(h)+(i)$.

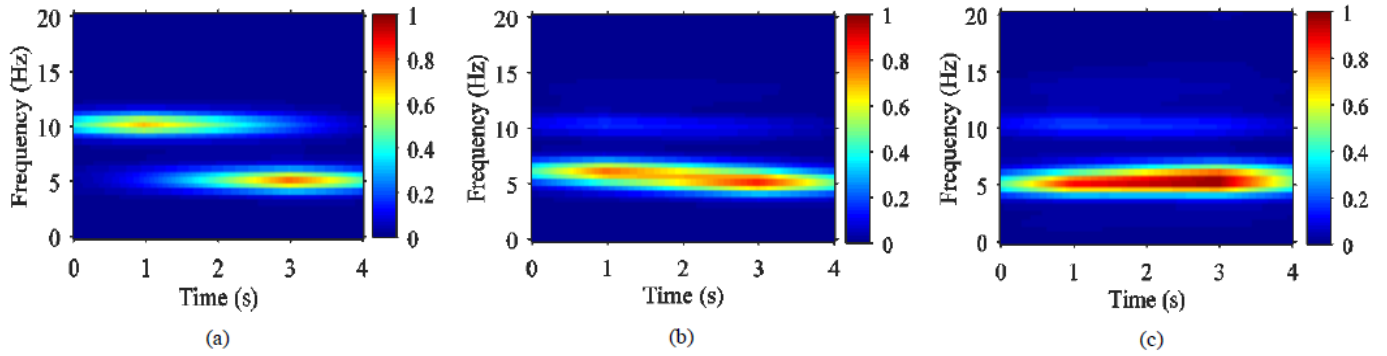


Fig. 4. Artificial signal STFT-dependent TF consistency among (a) X and Y, (b) X and Z, and (c) Y and Z.

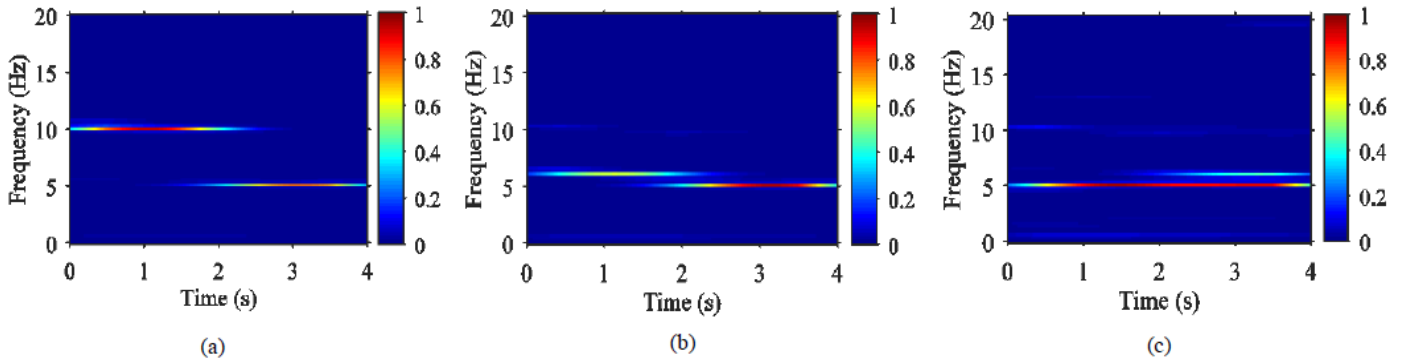


Fig. 5. Artificial signal coherence (a) between X as well as Y, (b) between X and Z, and (c) between Y and Z using SST-dependent TF.

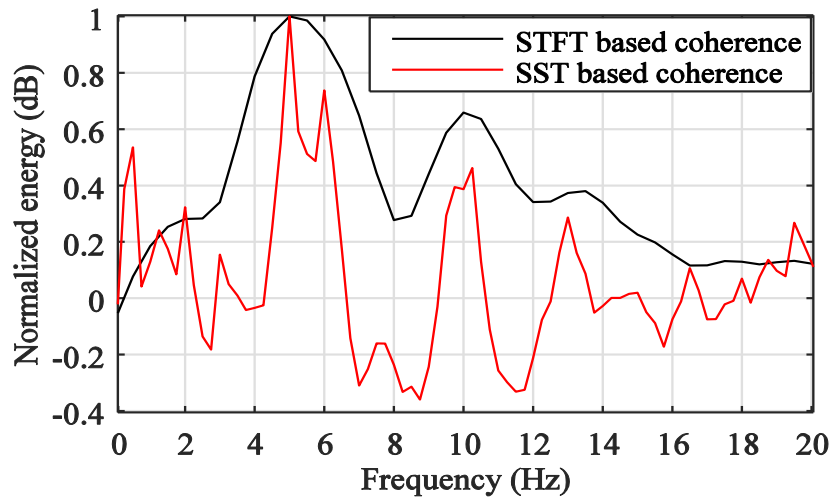


Fig. 6. Marginal frequency consistencies of STFT-depended coherence (black line) with SST-depended coherence (red line) between the artificial signals Y and Z.

2) *Real data:* The actual EEG data was gathered from the calib_ds1a. (IV dataset) generated from the 4th Brain-Computer Interface (BCI) Competition, which is openly accessible. The information is used to determine how well the suggested strategy performs. It is noted in subjects who are in good health. Motor imagery is accomplished during the entire session without any input. Two kinds of motor images are chosen from left hand, right hand, and foot movement for each

subject. Movement signals of the left hand and foot are present continuously in the calibration dataset calib_ds1a. 59 EEG channels comprising 200 trials lasting four seconds each make up the data. The data are sampled at a rate of 100 Hz. The data offset from the EEG signals has been eliminated during pre-processing. In order to get the alpha frequency band, which has intricate patterns of intermittent synchronization, The brain wave then goes across a 4th-order

Butterworth band transfer filter with a frequency range of 8 Hz to 12 Hz. [8]. The inter-channel coherence is measured in this experiment using the two channels T7 and T8. Fig. 7 shows the unprocessed EEG signal, the purified alpha ingredient, as well as the spectrum of alpha for the left-hand movement channels T7 and T8. The foot movement data from channels T7 and T8 are similarly depicted in Fig. 8. The STFT-based time-frequency coherence for left hand and foot movement motor images are shown in Fig. 9(a) and 9(c) for channels T7 and T8. The time-frequency coherence between channels T7 and T8 of left hand and foot movement, based on SST, are shown in Fig. 9(b) as well as 9(d), correspondingly. Fig. 9 shows how SST-based TF coherence, in contrast to STFT, exhibits remarkable localization of extremely small band frequency components.

3) *BCI interpretation:* In this investigation, the time-frequency coherence across channels in the left and right hemispheres of the human brain, is investigated. Moreover, the distinction involving left hand and left foot action in motor imagery is seen. Sensorimotor rhythms can be managed with the help of motor imagery [22], and the patterns are more active in the central region of the brain [23]. The 59 EEG channels are therefore divided into three channels from each hemisphere (T7, FC5, and CP5) and three channels from each hemisphere (T8, FC6, and CP6) for coherence assessment. Fig. 11 depicts the spatial distribution of the scalp's channels in the 10/20 EEG system. To evaluate time-frequency coherences, there are eight-channel clusters used: FC5FC6, FC5T8,

FC5CP6, T7FC6, T7T8, T7CP6, CP5FC6, CP5T8, as well as CP5CP6. On every one of the chosen channel pairings, time-frequency coherences based on STFT and SST are assessed. Eq. (9) determines how to weight the time-frequency coherences

$$\left|C_{x,y}(t, f)\right|_{weighted}^2 = \left|C_{x,y}(t, f)\right|^2 \bullet \tilde{C}_{x,y}(f) \quad (9)$$

Here, the weight matrix is the marginal frequency coherence and the notation \bullet is a binary singleton multiplication function. Using the weighted time-frequency coherence, the marginal time coherence is computed. It is said that minimal time coherence is

$$\tilde{C}_{x,y}(t) = \arg \max_f \left(\left|C_{x,y}(t, f)\right|_{weighted}^2 \right); t = 1, 2, 3, \dots, T \quad (10)$$

The minimal time consistency in this study is determined by averaging the marginal time coherence throughout 100 trials. The normalized readings during the time for several connection pairs of left-hand and foot swing data are displayed in Fig. 12. Data on left-hand movement is represented by solid lines, and information on foot movement is represented by dashed lines. Fig. 12's left panel displays SST-based marginal temporal coherences, while the right panel displays STFT-dependent marginal time consistencies. For both the left hand as well as foot activities sensory motor imaging information are distinguishable using the SST-based marginal time coherence.

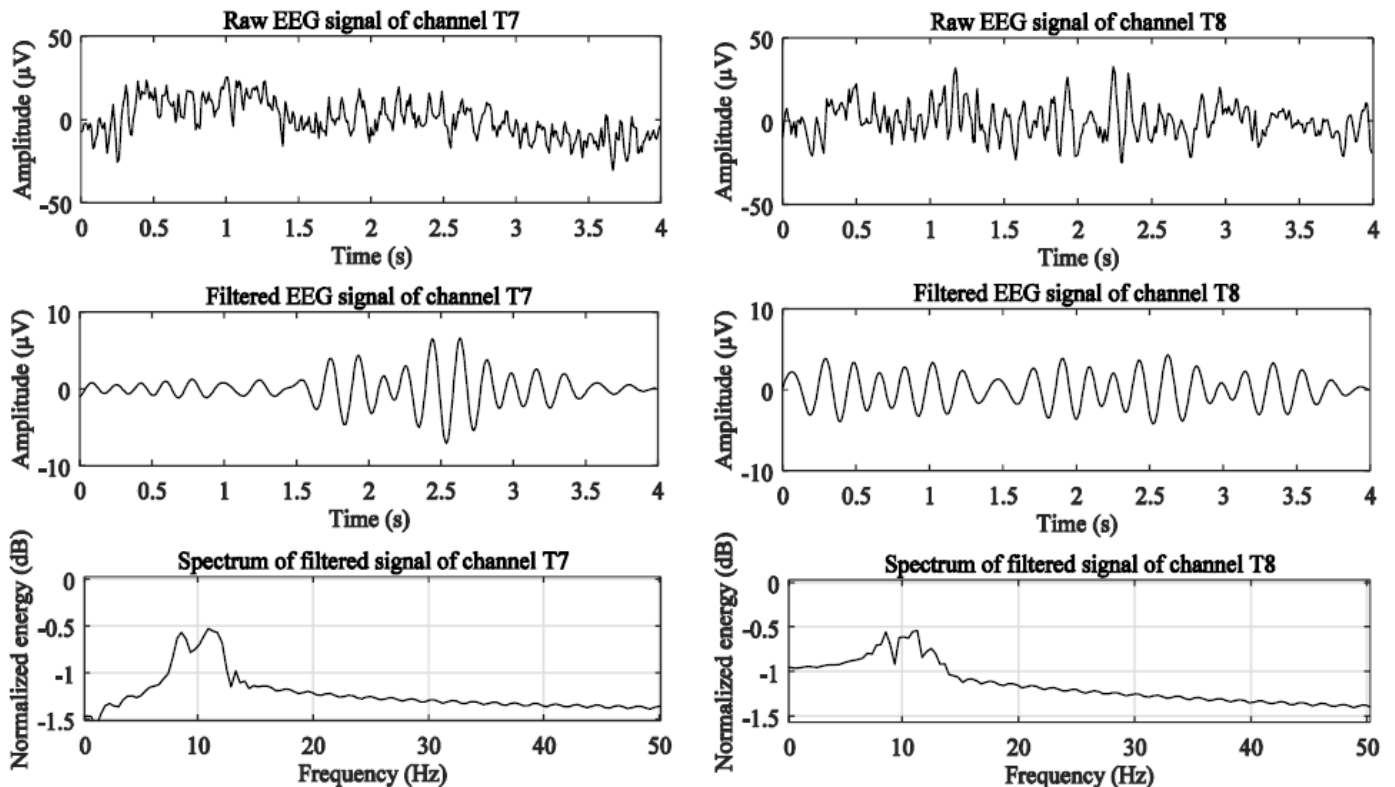


Fig. 7. Left hand movement data: first row is the raw EEG signals, second and third row are the filtered EEG signals and spectrums of the filtered component respectively.

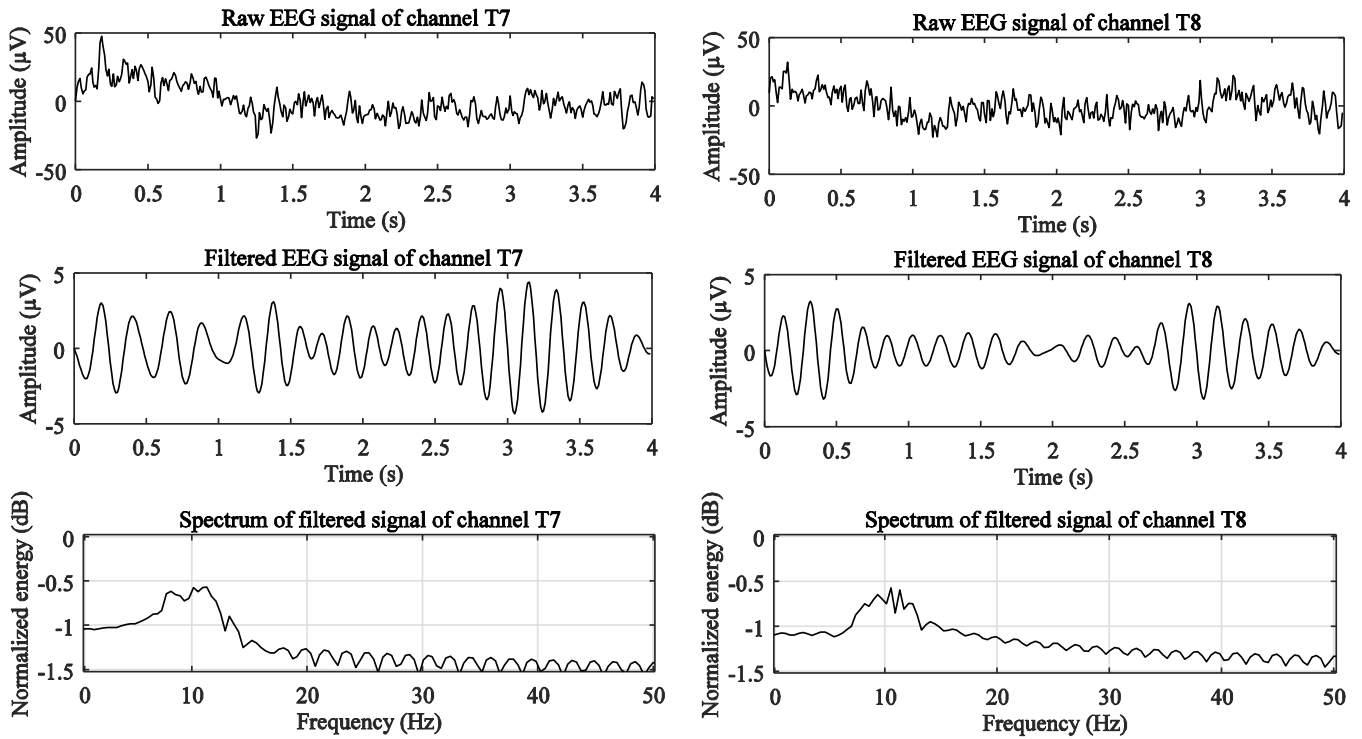


Fig. 8. Foot movement data: first row is the raw EEG signals, second and third row are the filtered EEG signals and spectrums of the filtered component respectively.

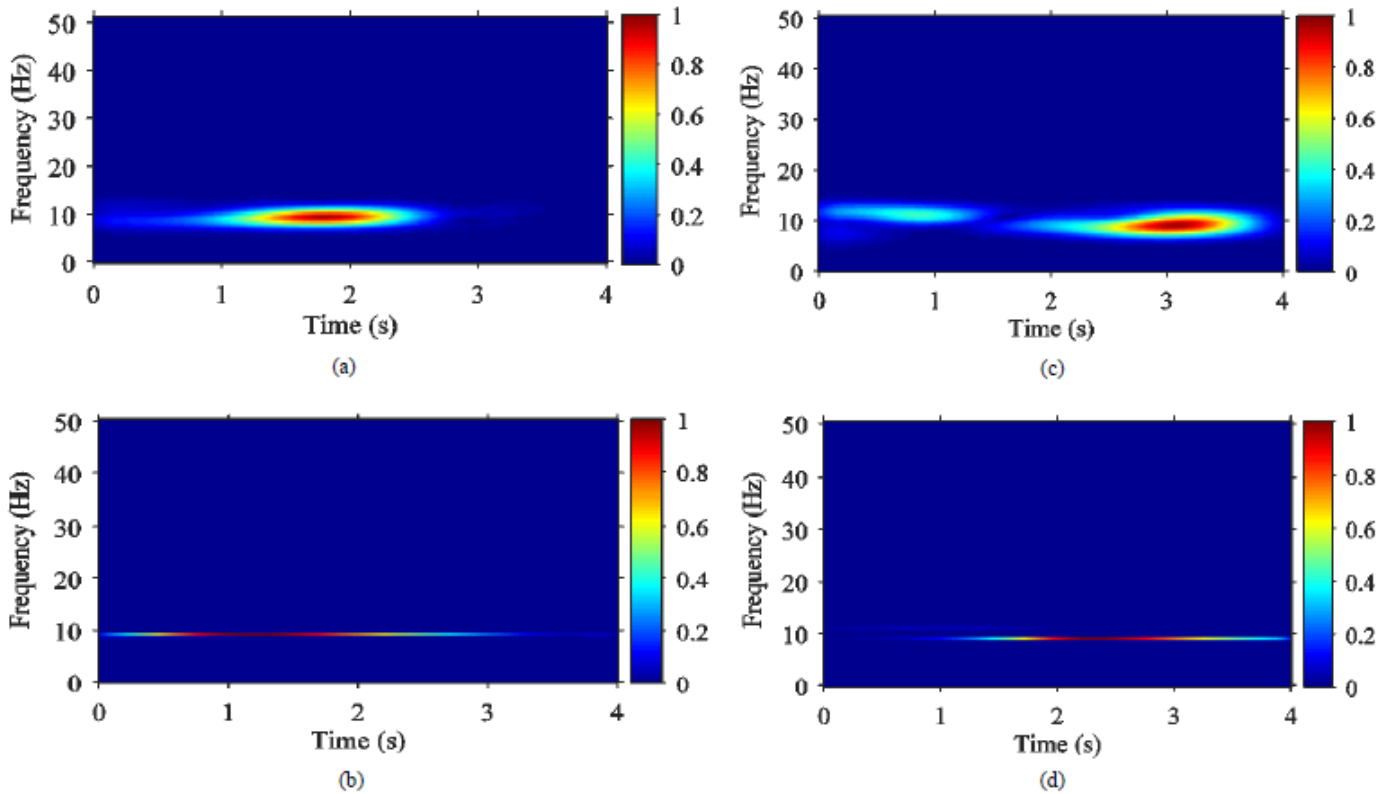


Fig. 9. TF coherence between channels T7 and T8 based on (a) STFT and (b) SST of left hand movement data, (c) STFT and (d) SST of foot movement data.

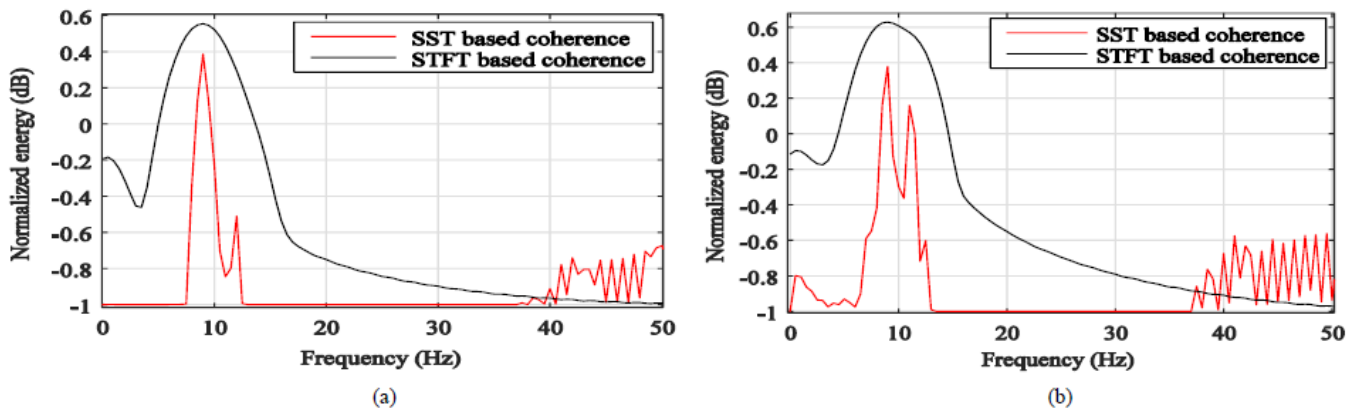


Fig. 10. Left hand action (a) as well as foot action (b) of channels T7 and T8 with marginal frequency consistencies.

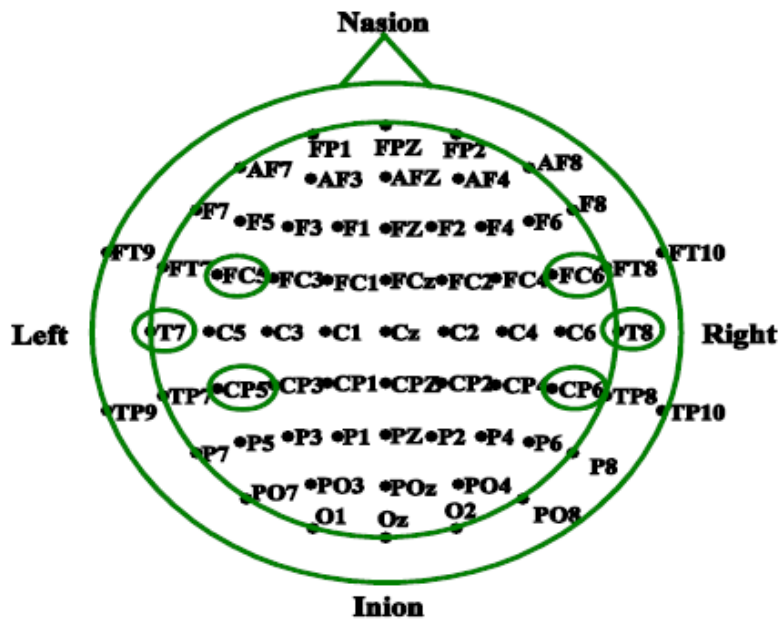
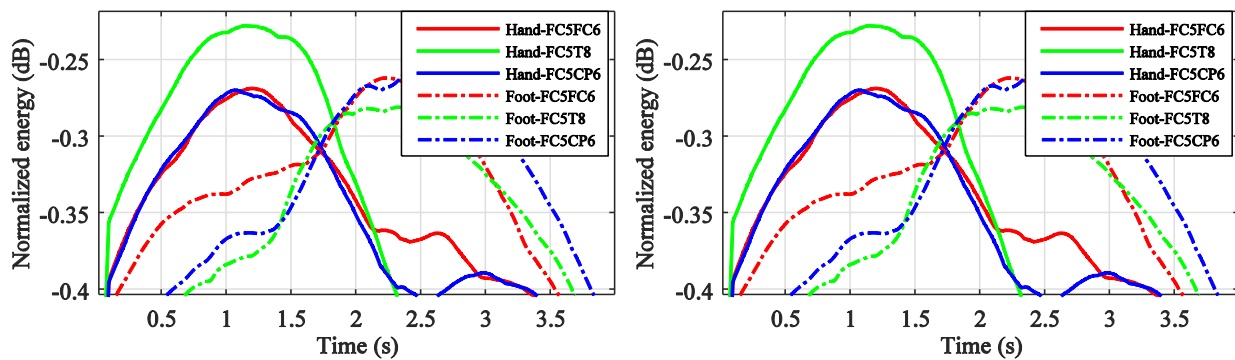


Fig. 11. The American EEG Society has standardized the electrode map of something like the 10/20 EEG system. For the dataset utilized in this experiment, the marked conductors T7, FC5, as well as CP5 beginning the left cerebral hemisphere and T8, FC6, as well as CP6 as of the right side of the brain were chosen.



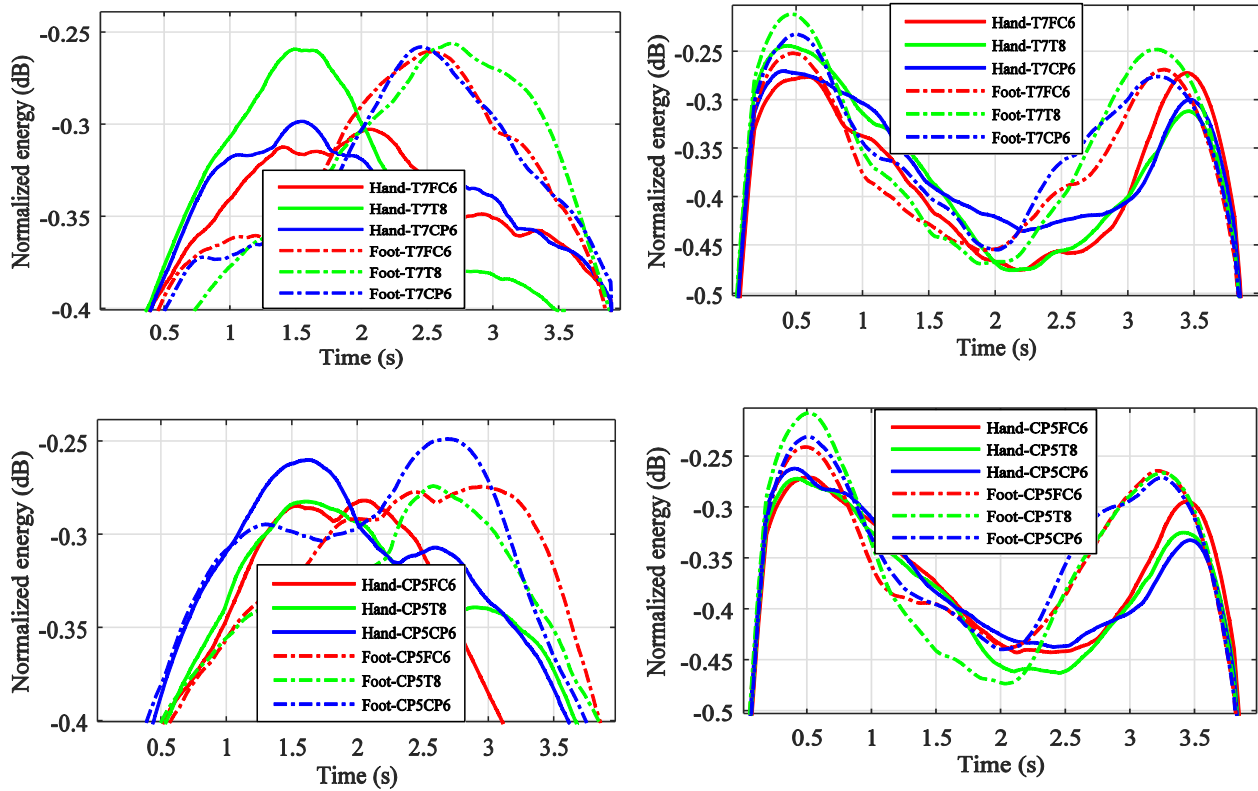


Fig. 12. Movement information for the left hand and foot for STFT-dependent (right panel) and SST-based models (left panel) approaches exhibit a small amount of time coherence between distinct channel pairings.

IV. DISCUSSION

EEG outputs are used in the research to examine how well SST performs in time-frequency illustration. Fig. 9 displays the TFR derived from left hand and foot movement data utilizing STFT and SST motor images. Fig. 10 shows, for left hand and left foot movement data, the energy related to the marginal frequency consistency. Although the SST-based technique shows acute localization of each frequency element within a relatively small band of frequencies, the marginal frequency coherence depended on STFT displays weak localization of frequency agents. The frequencies of 9 Hz and 11 Hz in Fig. 10(b) can be clearly distinguished using SST-based marginal frequency coherence, whereas they cannot be achieved using an STFT-dependent method. From now, SST-dependent time-frequency consistency is better than STFT-based time-frequency coherence. The fundamental cause is that the employment of something like a window function for covering in STFT results in the introduction of cross-spectral energy, which causes the energy to spread over a broad range of frequencies. The TFR performance of the SST has been evaluated in our previous work [24]. This paper is a development of our earlier paper [24]. Our prior work is expanded upon in this one. In addition, the smoothing operations on the TFR, as well as BCI interpretation, are introduced in this paper. BCI can use marginal frequency coherence followed by marginal time coherence based on SST. The coherence value for left-hand movement data in the left panel of Fig. 12 is at its highest in the time range of 1-2 seconds, whereas the coherence value for foot movement data

across all channel pairs is at its highest in the time range of two to three seconds. In the SST-based marginal temporal coherence, left-hand and foot measurement data are explicitly distinguished from one another. The marginal temporal coherence model based on STFT, however, does not exhibit this kind of selectivity (in the right panel). The main cause is the fact that the STFT has a fixed time-frequency window while the SST has a changeable one, making it difficult to accurately evaluate signals with broad bandwidths that fluctuate rapidly over time. Moreover, the STFT demands that the brain wave be stationary for a given time period, yet EEG signals exhibit non-stationary characteristics.

V. CONCLUSIONS

The analysis of the time-frequency (TF) consistency among two signals is offered in this work using an innovative approach. For each of the provided signals, the time-frequency densities of the crossed and auto spectrums are calculated. Then, using quasi smoothing agents, the spectral densities are smoothed. The TF coherence is calculated using smooth spectral densities for artificial signals with time-frequency representations based on STFT and SST. A genuine EEG signal with various motor images is used to test the suggested SST-based coherence estimate method. Comparing the two strategies' performances reveals that the SST-based approach is more effective than STFT at locating frequency contents with greater spatial precision. Then, using both SST and STFT-based coherences, marginal time coherences are computed. It is clearly shown that the STFT-dependent

marginal time consistencies are incapable to distinguish among left hand and foot activity data, in contrast to the SST-dependent marginal time coherences, which can. This implies that these marginal time coherences can enhance BCI design. In order to get greater performance, it is advised BCI designers to take these coherences as supplementary features when designing a BCI system.

Future study might explore brand-new combinations of features and feature selection, as well as the use of these features for BCI tasks other than motor imagery. Additionally, there is a need for work in the design of novel algorithms, including physiologically realistic error functions for EEG signal predictions for the complexity feature.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper

REFERENCES

- [1] Gregoriou GG, Gots SJ, Zhou H, Desimone R (2009) High-Frequency, long-range coupling between prefrontal and visual cortex during attention. *Sci.* 324:1207–1210.
- [2] Liang H, Bressler SL, Ding M, Desimone R, Fries P (2003) Temporal dynamics of attention-modulated neuronal synchronization in macaque V4. *Neurocomputing* 52:481–487.
- [3] Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2003) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc. Natl. Acad. Sci. U.S.A.* 101:9849–9854.
- [4] Moller E, Schack B, Arnold M, Witte H (2001) Instantaneous multivariate EEG coherence analysis by means of adaptive high-dimensional autoregressive models. *J. Neurosci. Methods* 105:143–158.
- [5] Daubechies I, Lu J, Wu H-T (2011) Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. *Appl. Computational Harmonic Anal.* 30:243–261.
- [6] Ahrabian A, Looney D, Stankovic L, Mandic DP (2015) Synchrosqueezing-based time-frequency analysis of multivariate data. *Signal Process.* 106:331–341.
- [7] Auger F, Flandrin P, Lin Y-T, McLaughlin S, Meignen S, Oberlin T, Wu H-T (2013) Time-frequency reassignment and synchrosqueezing: an overview. *IEEE Signal Process. Mag.* 30: 32–41.
- [8] Mehrkanoon S, Breakspear M, Daffertshofer A, Boonstra TW (2013) Non-identical smoothing operators for estimating time-frequency interdependence in electrophysiological recordings. *EURASIP J. Advances Signal Process.* 2013:1–16.
- [9] Cohen E, Walden A (2010) A statistical study of temporally smoothed wavelet coherence. *IEEE Trans. Signal Process.* 58:2964–2973.
- [10] Jin, J., Liu, C., Daly, I., Miao, Y., Li, S., Wang, X., & Cichocki, A. (2020). Bispectrum-based channel selection for motor imagery-based brain-computer interfacing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(10), 2153–2163.
- [11] Suemitsu, K., & Nambu, I. (2023). Effects of Data Including Visual Presentation and Rest Time on Classification of Motor Imagery of Using Brain-Computer Interface Competition Datasets. *IEEE Access*. Jun 12
- [12] Molla, M. K. I., Al Shiam, A., Islam, M. R., & Tanaka, T. (2020). Discriminative feature selection-based motor imagery classification using EEG signal. *IEEE Access*, 8, 98255–98265.
- [13] Das, K., & Pachori, R. B. (2022). Electroencephalogram based motor imagery brain computer interface using multivariate iterative filtering and spatial filtering. *IEEE Transactions on Cognitive and Developmental Systems*.
- [14] Li C, Liang M (2012) A generalized synchrosqueezing transform for enhancing signal time-frequency representation. *Signal Process.* 92: 2264–2274.
- [15] Wu W-T, Flandrin P, Daubechies I (2011) One or two frequencies? The synchrosqueezing answers. *Advances Adaptive Data Anal.* 3:29–39.
- [16] Tiesinga P, Fellous J-M, Sejnowski TJ (2008) Regulation of spike timing in visual cortical circuits. *Nat. Rev. Neurosci.* 9:97–107.
- [17] Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2:539–550.
- [18] Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* 32: 209–224.
- [19] Lowet E, Roberts MJ, Bonizzi P, Karel J, De Weerd P (2016) Quantifying neural oscillatory synchronization: a comparison between spectral coherence and phase-locking value approaches. *PLoS one* 11.
- [20] Zhang ZG, Cai XL, Chan SC, Hu Y, Hu L, Chang CQ (2009) Time-frequency coherence analysis of multi-channel event-related potential using adaptive windowed Lomb periodogram. *4th International IEEE EMBS Conf. Neural Eng.* 657–660.
- [21] Brittain JS, Halliday DM, Conway BA, Nielsen JB (2007) Single-trial multiwavelet coherence in application to neurophysiological time series. *IEEE Trans. Biomed. Eng.* 54:854–862.
- [22] McCreddie KA, Coyle DH, Prasad G (2013) Sensorimotor learning with stereo auditory feedback for a brain-computer interface. *Med Biol Eng Comput* 51: 285–293.
- [23] Hadjidimitriou S, Zacharakis A, Doulgeris P, Panoulas K, Hadjileontiadis L, Panas S (2010) Sensorimotor cortical response during motion reflecting audiovisual stimulation: evidence from fractal EEG analysis. *Med Biol Eng Comput* 48: 561–572.
- [24] Ali MS, Ferdous MJ, Hamid ME, Molla MKI, (2016) Time-frequency coherence of multichannel EEG signals: synchrosqueezing transform based analysis. *International Journal of Computer Science Trends and Technology (IJCTST)* 4:40–48.
- [25] Bang, J. S., Lee, M. H., Fazli, S., Guan, C., & Lee, S. W. (2021). Spatio-spectral feature representation for motor imagery classification using convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7), 3038–3049.
- [26] Saranyasoontorn K, Manuel L, Veers PS (2004) A comparison of standard coherence models for inflow turbulence with estimates from field measurements. *J. Sol. Energy Eng.* 126:1069–1082.
- [27] Shovon, T. H., Al Nazi, Z., Dash, S., & Hossain, M. F. (2019, September). Classification of motor imagery EEG signals with multi-input convolutional neural network by augmenting STFT. In *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)* (pp. 398–403). IEEE.
- [28] Xu, B., Zhang, L., Song, A., Wu, C., Li, W., Zhang, D., ... & Zeng, H. (2018). Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *Ieee Access*, 7, 6084–6093.

Systematic Review for Phonocardiography Classification Based on Machine Learning

Abdullah Altaf¹, Hairulnizam Mahdin², Awais Mahmood³, Mohd Izuan Hafez Ninggal⁴,
Abdulrehman Altaf⁵, Irfan Javid⁶

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia,
Parit Raja, Batu Pahat, Johor, Malaysia^{1, 2, 5}

Computer Engineering Dept, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia³

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Selangor, Malaysia⁴

Department of Computer Science & IT, University of Poonch Rawalakot, AJK Pakistan⁶

Abstract—Phonocardiography, the recording and analysis of heart sounds, has become an essential tool in diagnosing cardiovascular diseases (CVDs). In recent years, machine learning and deep learning techniques have dramatically improved the automation of phonocardiogram classification, making it possible to delve deeper into intricate patterns that were previously difficult to discern. Deep learning, in particular, leverages layered neural networks to process data in complex ways, mimicking how the human brain works. This has contributed to more accurate and efficient diagnoses. This systematic review aims to examine the existing literature on phonocardiography classification based on machine learning, focusing on algorithms, datasets, feature extraction methods, and classification models utilized. The materials and methods used in the study involve a comprehensive search of relevant literature and a critical evaluation of the selected studies. The review also discusses the challenges encountered in this field, especially when incorporating deep learning techniques, and suggests future research directions. Key findings indicate the potential of machine and deep learning in enhancing the accuracy of phonocardiography classification, thereby improving cardiovascular disease diagnosis and patient care. The study concludes by summarizing the overall implications and recommendations for further advancements in this area.

Keywords—Heart sounds classification; Phonocardiogram (PCG); CVDs; deep learning

I. INTRODUCTION

Phonocardiography (PCG) is one of the basic techniques used to understand the heart's state and assess whether the heart is in a natural state or has some abnormal pattern. PCG is a diagnostic procedure that allows a visual record of the sounds and murmur created by the contracting heart, including its valves and connected large vessels. In the absence of diagnosis equipment, the stethoscope is only the tool available to general physicians to examine a patient's heart sounds [1]. Cardio-specialist can understand the heartbeat as a specialist and recommend further medical procedures to the patients according to their heart condition but usually, in the unavailability of a cardio-specialist, the general physicians cannot detect, if the heart is functioning properly or if there is any type of exception due to the closure of heart walls. Environmental interferences, such as those caused by friction between the device and a human's skin, Electromagnetic

Interference (EI), and unrelated noises like breath, lung, and ambient sounds, can readily interfere with the process of PCG because signals in the form of sound generated by the human heart are frequently paired with EI, out-of-band noise must be removed [2], [3].

To cope with the limitations in traditional phonocardiography techniques, machine learning (ML) based methods can be a good solution for phonocardiography for several reasons [4]. ML models can automate the process of analyzing PCG recordings, which can be time-consuming and subject to human error when done manually. ML models can be trained on large datasets of labeled PCG recordings, which can improve their accuracy in detecting and diagnosing heart conditions. ML-based methods can handle large amounts of data, which is important in PCG as it requires analyzing audio signals over time and adapting to new data, and improving their performance over time, which can be useful in handling diverse populations and detecting new conditions [5].

Using ML-based approaches to achieve real-time heart disease detection from audio signals is challenging because heart sounds can vary significantly depending on factors such as the person's age, sex, and underlying medical conditions [6]. This makes it difficult to develop an ML model that can accurately detect and classify heart sounds in a wide range of individuals. Heart sounds can be difficult to distinguish from other sounds in the body, such as breathing and blood flow [7]. Additionally, external noise such as background noise or equipment noise can also interfere with the recording and analysis of heart sounds [5]. Collecting a large and diverse dataset of heart sounds for training machine learning models can be difficult and time-consuming.

Heart sounds are complex and have a lot of variations, which can make it difficult for machine learning algorithms to accurately classify them. Overfitting is a common problem when building machine learning models, and can occur when a model is trained on a limited dataset and then performs poorly on new, unseen data [8]. The interpretation of phonocardiography signals requires expertise and knowledge of human anatomy and physiology, this is a challenge when using machine learning algorithms to interpret the signals. Designing a phonocardiography system using machine learning is a big challenge because it requires overcoming several

technical and logistical hurdles in accurately and reliably detecting a classifying heart sound [9]. The core phases involved in the phonocardiography process i.e., segmentation, feature extraction, and final classification results, are all considerably impacted by denoising. Preprocessing is required in phonocardiography systems to clean and enhance the quality of the raw phonocardiography signal before it is further analyzed.

It is necessary to create these components, incorporating the applicable criteria that follow.

II. AN OVERVIEW OF PHONOCARDIOGRAPHY

Machine Learning based phonocardiography systems models used in real-time to analyze Phonocardiography (PCG) recordings and provide diagnostic information, which can be useful in critical care settings. Overall, the combination of machine learning and PCG data can provide more accurate and objective results and helps to improve the diagnosis and monitoring of heart conditions [10]. The generic ML-based phonocardiography systems are depicted in Fig. 1.

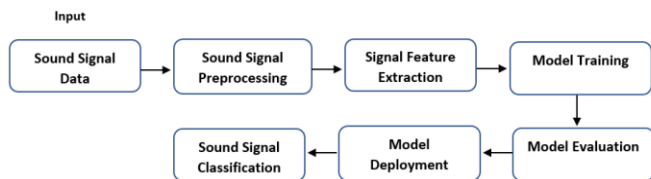


Fig. 1. ML-based phonocardiography general framework.

A. Heart Sound Signal

Heart sound signals, also known as phonocardiograms or PCG signals, are acoustic signals generated by the mechanical activities of the heart. These signals provide important information about the structure, function, and abnormalities of the cardiovascular system. Heart sound signals typically consist of two prominent components: the first heart sound (S1) and the second heart sound (S2). S1 is produced by the closure of the mitral and tricuspid valves during the systolic phase of the cardiac cycle, while S2 is generated by the closure of the aortic and pulmonary valves during the diastolic phase. These components are accompanied by additional sounds such as the third heart sound (S3) and fourth heart sound (S4), which can indicate specific cardiac conditions.

Heart sound signals are characterized by their duration, intensity, frequency content, and temporal relationships between different components. They contain valuable information about heart rate, heart rhythm, valve function, and the presence of murmurs, stenosis, or other cardiovascular abnormalities. Traditionally, heart sound signals were recorded using specialized electronic stethoscopes or phonocardiography equipment. However, with advancements in technology, heart sound signals can now be captured using digital stethoscopes, wearable devices, or even smartphone applications equipped with appropriate sensors.

Heart sounds (signals) are produced from a specific cardiac event such as the closure of a valve or tensing of a chordae tendineae. Most normal heart sound (Lub, Dub) signal rates at rest are between about 60 and 100 beats per minute. Sound is the pressure of air propagating to our ears. The digital audio

file is gotten from a sound sensor that can detect sound waves and convert them to electrical signals.

B. Normal Heart Sound Signal

Normal heart sounds, also known as physiological heart sounds, are the characteristic sounds produced by a healthy heart during its regular functioning. These sounds are a result of the synchronized mechanical activities of the heart's valves and chambers. The normal heart sound consists of two primary components: the first heart sound (S1) and the second heart sound (S2). S1 is a low-frequency sound that occurs at the beginning of each cardiac cycle and is caused by the closure of the mitral and tricuspid valves. S2 is a higher-pitched sound that occurs at the end of the cardiac cycle and is produced by the closure of the aortic and pulmonary valves. These two sounds create the familiar "lub-dub" rhythm associated with a normal heartbeat. The normal heart sound signifies the proper functioning of the heart's valves and chambers, reflecting a healthy cardiovascular system. Understanding the characteristics and timing of normal heart sounds is crucial in differentiating them from abnormal sounds and diagnosing various cardiac conditions [11].

C. Artifact Heart Sound Signal

Artifact heart sounds refer to extraneous or spurious sounds that may be inadvertently recorded or introduced during the process of capturing heart sound signals. These sounds are not of physiological origin and do not reflect the actual functioning of the heart. Artifact heart sounds can arise from various sources, such as environmental noise, patient movement, electrical interference, or improper placement of the recording device. They can manifest as random noise, clicking sounds, buzzing, or other irregular patterns that may obscure or distort the true heart sounds. Artifact heart sounds pose a challenge in the accurate analysis and interpretation of heart sound signals, as they can interfere with the detection of abnormal cardiac conditions or mask important diagnostic information. Efforts are made to minimize artifacts during data collection by ensuring proper recording techniques, reducing environmental noise and employing noise cancellation methods. Additionally, careful signal processing and expert interpretation are essential to distinguish artifact heart sounds from genuine physiological sounds and ensure the reliability and accuracy of heart sound analysis in clinical practice and research [12].

D. Extrastole Heart Sound Signal

Extrastole, also known as an extra heart sound or premature beat, refers to an abnormal additional sound that occurs in the cardiac cycle, occurring either before or after the normal heart sounds. It is typically characterized by a distinctive "gallop" or "clicking" sound. Extrastole is caused by premature contractions of the heart's ventricles, atria, or both. These premature contractions disrupt the normal rhythm and timing of the cardiac cycle. Common types of Extrastole include atrial premature complexes (APCs) and ventricular premature complexes (VPCs). Extrastole can be indicative of underlying heart conditions such as arrhythmias, valvular disorders, or heart muscle abnormalities. Detecting and analyzing Extrastole in heart sound signals is crucial for diagnosing and monitoring cardiac abnormalities. Advanced signal processing techniques and machine learning algorithms are employed to identify and

classify Extrastole patterns accurately. Understanding the presence and characteristics of Extrastole heart sounds aids in the comprehensive evaluation and management of cardiovascular health [13].

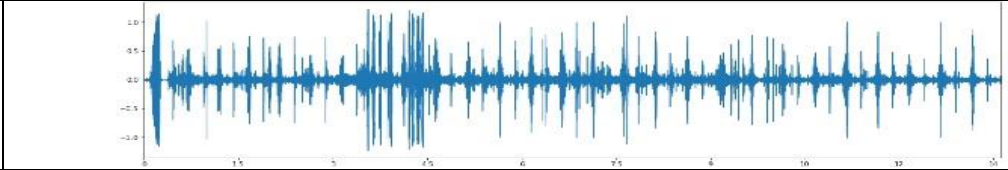
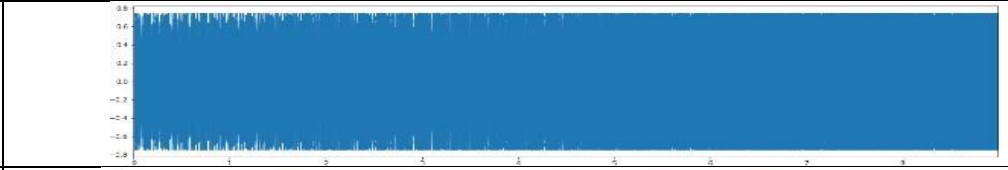
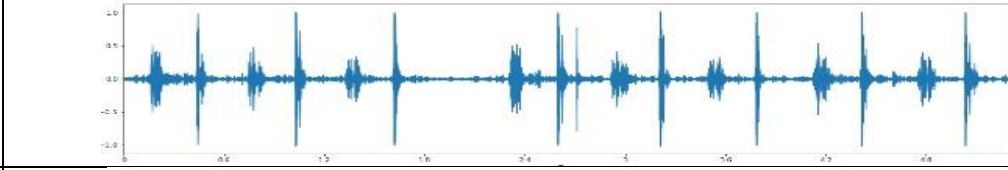

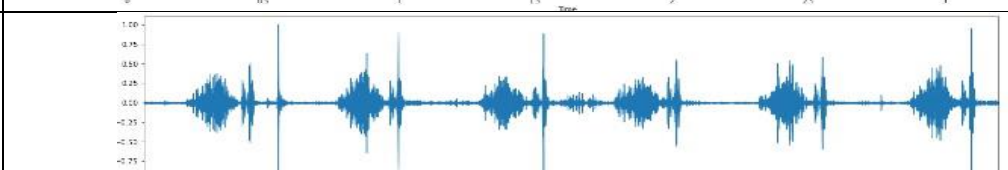
E. Extrahls Heart Sound Signal

Extrahls, also known as extraneous heart sounds or adventitious heart sounds, refer to abnormal sounds that are superimposed on normal heart sound signals. These sounds are not typically associated with the regular functioning of the heart and can arise from various pathological conditions or abnormalities within the cardiovascular system. Extrahls can manifest as additional clicks, murmurs, or abnormal sounds that occur in addition to the first and second heart sounds. They can be indicative of structural heart defects, valve abnormalities, turbulent blood flow, or other cardiac disorders. Analyzing Extrahls in heart sound signals is crucial for diagnosing and monitoring cardiovascular conditions, as they can provide valuable insights into the presence and severity of cardiac abnormalities. Advanced signal processing techniques, such as spectrogram analysis, wavelet analysis, or machine learning algorithms, are employed to identify and characterize Extrahls accurately. By detecting and analyzing Extrahls in heart sound signals, clinicians and researchers can improve their understanding of cardiac pathologies and make informed decisions regarding patient care and treatment strategies [14].

F. Murmur Heart Sound Signal

Murmur heart sound signals refer to abnormal or atypical sounds that are heard during auscultation of the heart. Murmurs are characterized by a prolonged, swishing, or whooshing sound that occurs between the normal heart sounds (S1 and S2). These sounds are caused by turbulent blood flow within the heart or blood vessels, typically due to structural abnormalities such as valve defects, stenosis, regurgitation, or abnormal blood flow patterns. Murmurs can be classified based on their timing, intensity, pitch, and location within the cardiac cycle. They are often graded on a scale from 1 to 6, with higher grades indicating more pronounced murmurs. Accurate identification and characterization of murmurs in heart sound signals are crucial for diagnosing and managing various cardiovascular conditions. Advanced signal processing techniques, such as spectral analysis, time-frequency analysis, or machine learning algorithms, can aid in the automated detection and classification of murmur patterns. By analyzing murmurs in heart sound signals, healthcare professionals can assess the severity of underlying cardiac abnormalities, determine appropriate treatment strategies, and monitor the effectiveness of interventions for improved patient care [12]. Table I depicts a human’s heart-generated sound Wave-form signal.

TABLE I. HEART SOUND WAVE-FORM SIGNAL, X-AXIS REPRESENTS TIME AND Y-AXIS MEASURES AMPLITUDE (GENERATED BY USING PYTHON LIBRARY: MATPLOTLIB)

S.No	Heart State	Human’s Heart Generated Sound Wave-form Signal
1.	Normal	
2.	Artifact	
3.	Extrastole	
4.	Extrahls	
5.	Murmur	

G. Signal Denoising

Signal denoising refers to the process of removing unwanted noise or interference from a signal while preserving the underlying information of interest. It is a fundamental technique in signal processing used to improve the quality and reliability of signals in various domains, such as audio, image, and biomedical signals. The presence of noise in a signal can distort or mask important features, making it challenging to extract meaningful information or make accurate measurements. Signal denoising methods aim to reduce or eliminate this noise, enhancing the signal's clarity and fidelity. These methods employ various techniques, such as filtering, statistical analysis, wavelet transforms, or machine learning algorithms, to suppress or attenuate the noise components while preserving the desired signal components. Signal denoising is widely used in applications where the accuracy and reliability of signal analysis, interpretation, or decision-making are critical, allowing researchers, engineers, and practitioners to obtain cleaner and more accurate signals for further analysis or processing.

Signal denoising is highly important in various applications due to its ability to enhance signal quality, improve data analysis accuracy, facilitate signal interpretation, increase measurement precision, optimize signal processing techniques, enable better signal visualization, and enhance communication and signal transmission reliability. By removing unwanted noise and interference from signals, signal denoising improves the accuracy, clarity, and reliability of the underlying information, leading to more meaningful analysis, interpretation, and decision-making. It is a crucial step in fields such as biomedical signal processing, image and audio processing, communication systems, and scientific research, where accurate and reliable signal analysis is essential for successful outcomes.

Environmental interferences, such as those caused by friction between the device and a human's skin, Electromagnetic Interference (EI), and unrelated noises like breath, lung, and ambient sounds, can readily interfere with the process of recording heart sounds [15]. Because signals in the form of sound generated by the human heart are frequently paired with EI, out-of-band noise must be removed. The segmentation, feature extraction, and final classification results are all considerably impacted by denoising. Wavelet denoising, variational mode deconstruction denoising, and Digital Filter Denoising (DFD) are the three most often used denoising techniques [16]. A new line of study in the field of heart sound feature extraction is the creation of a wavelet function for the human heart's signals based on the past understanding of heart sound data [17].

A sound spectrum refers to the distribution of the different frequencies present in a sound signal. It represents the energy or intensity of each frequency component within the signal. The spectrum provides valuable information about the composition and characteristics of the sound, allowing for the identification and analysis of specific frequency components. In the context of heart sound signals, a sound spectrum can reveal the presence and intensity of different sound frequencies

associated with normal or abnormal heart sounds. By analyzing the spectrum, healthcare professionals and researchers can gain insights into the underlying physiological conditions and abnormalities of the heart.

Spectral analysis techniques, such as Fourier transform or wavelet transform, are commonly employed to compute the sound spectrum and visualize the frequency content of the signal. This information can assist in the diagnosis, monitoring, and treatment of various cardiovascular disorders, providing valuable insights into the acoustic properties of the heart. Table II depicts different Heart Sound spectrums after denoising of heart's sound wave signal.


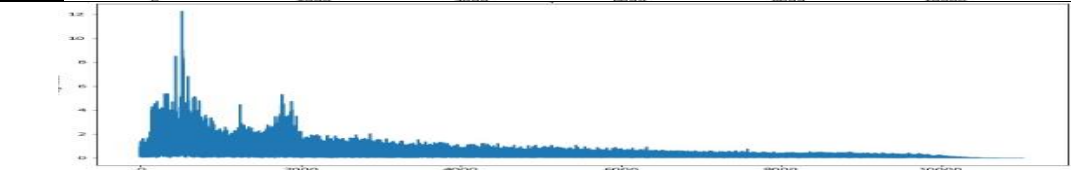
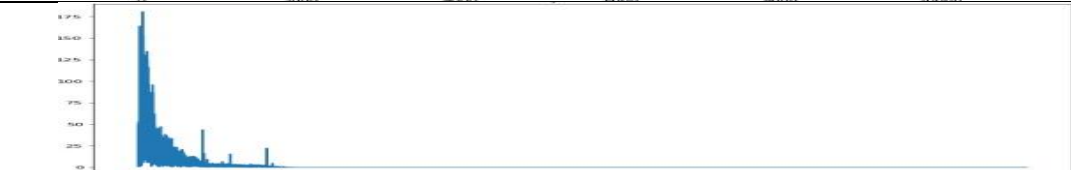


H. Signal Segmentation

Signal segmentation is a fundamental process in signal processing that involves dividing a continuous signal into shorter segments or intervals. This technique is essential for isolating specific regions of interest within a signal, allowing for targeted analysis and processing. Signal segmentation is commonly used in various fields, such as speech recognition, audio processing, image analysis, and biomedical signal analysis. By segmenting a signal, researchers can focus on specific time intervals or frequency components for further analysis, enabling the extraction of meaningful features and patterns. This approach facilitates tasks such as event detection, signal classification, anomaly detection, and time-frequency analysis. Additionally, signal segmentation helps in dealing with non-stationary signals by breaking them down into smaller, more manageable segments. Overall, signal segmentation plays a vital role in signal processing applications, allowing researchers and practitioners to effectively analyze and understand complex signals by examining their constituent segments individually.

As part of the segmentation process, the heart sounds of the first human (S1), the second human (S2), and the diastole are split into four parts or segments. Each section has useful components that help distinguish between the different types of heart sounds. However, individual differences in the length of the human heartbeat cycle, the number of human heart sounds, and the kinds of heart murmurs result in erroneous PCG signal segmentation. Thus, segmenting the FHS is a crucial step in the automated PCG analysis process.

In recent years, envelope-based techniques have been among the most frequently utilized techniques for segmenting heart sounds [18], [19]. Electrocardiogram (ECG) [20], feature-based methods [21], time-frequency analysis methods [22], and probabilistic model methods [23], [24], [25], [26] are some important segmentation methods. The underlying premise of the employed algorithms is that the diastolic interval is more prolonged than the systolic time. In actuality, especially in newborns and cardiac patients, this supposition is not always accurate for an aberrant heart sound [27]. Based on the similarity between the ECG and the human heart signals, it has been discovered that algorithms that combine the cardiac cycle with an ECG signal perform better at segmenting data. They do have higher hardware and software requirements, though.

TABLE II. HEART SOUND SPECTRUM, X-AXIS REPRESENTS FREQUENCY AND Y-AXIS REPRESENTS MAGNITUDE (GENERATED BY USING PYTHON LIBRARY: MATPLOTLIB)

S.No	Heart State	Human's Heart Sound Signal Spectrum
1.	Normal	
2.	Artifact	
3.	Extrastole	
4.	Extrahls	
5.	Murmur	

Humans can hear sound not only at a particular time by its intensity but also by its pitch. The pitch is the frequency of the sound, a higher pitch corresponds to a higher frequency, and vice versa. So, to have a representation that is closer to the human brain, another dimension which is frequency is added to the representation, which is the spectrogram.

III. REVIEW

This review will provides a compendious review of Phonocardiography, Machine Learning (ML) literature including experimental, empirical, and theoretical studies. The modern advancements and contributions from recent studies related to major and compelling administration depict this research.

Previously proposed phonocardiography, Machine learning techniques such as Deep Learning (DL), Extreme Learning Machines (ELMs), Deep Extreme Learning Machines (DELMS), and previous technologies and techniques are comprehensively reviewed. The literature mentioned is either the most benchmark research contributions, most relevant, highly cited, or published in well-reputed research journals.

A. Phonocardiography

Phonocardiography is a non-invasive diagnostic technique used to capture and analyze the sounds produced by the heart during its normal functioning. It involves recording the sounds made by the heart using a sensitive microphone or electronic sensor and then analyzing the recordings to identify any abnormal sounds or patterns that may indicate the presence of heart disease. Fig. 2 shows the Phonocardiography Signals of Normal Heart. In the phonocardiograph, the process of determining the state of the heart is done using the waves that come from the heart, and the process of classification of the heartbeats is done as normal and abnormal [28].

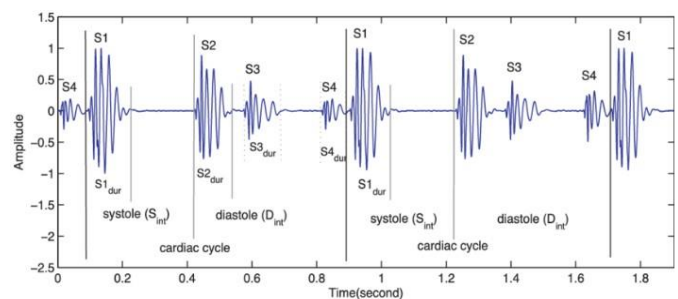


Fig. 2. Phonocardiography signals of normal heart.

The normal is defined as the state that an adult's blood pressure at rest varied from 60 to 100 beats per minute. In general, improved cardiac function and improved cardiovascular health are demonstrated by a lower resting heart rate. As for the abnormal state, the heart is in a state of disorder, unstable, and known any erratic heartbeat is known to be an abnormal heart rate or an increased heart rate. An irregular heart rhythm (rapid heartbeat, called tachyarrhythmia, or slow pulse, called slow arrhythmia-bradyarrhythmia) can accompany arrhythmia [29]. Fig. 3 shows the Phonocardiography Wave Signals of Arrhythmia-Bradyarrhythmia.

One of the most prominent application features of phonocardiography is the detection and monitoring of the evaluate the effectiveness of treatment of the key heart abnormalities symptoms like ventricular dysfunction, aortic regurgitation, mitral regurgitation, aortic stenosis, heart failure, hypertrophic cardiomyopathy, tricuspid regurgitation, coronary artery disease, pulmonic stenosis, atrial fibrillation, ventricular tachycardia, heart murmurs, aortic aneurysms, systolic heart murmurs, ventricular hypertrophy, ventricular septal defects and ductus arteriosus [30].

Left ventricular dysfunction refers to a condition where the heart's left ventricle, which is responsible for pumping oxygenated blood to the rest of the body, is not functioning properly. This can occur due to a variety of reasons, including damage to the heart muscle, high blood pressure, heart valve problems, or coronary artery disease. Left ventricular dysfunction can lead to a range of symptoms, including shortness of breath, fatigue, swelling in the legs and feet, and chest pain. It can also increase the risk of heart failure, heart attack, and other cardiovascular complications.

Phonocardiography-based detection of left ventricular dysfunction was reviewed in detail by Jagannath [31].

Some notable contributions along with the feature extraction techniques, datasets, modeling techniques, and results are depicted in Table III.

From Table III, it can be noted that most of the research work has been done using different feature extraction techniques including MFCC. Also, different researchers used different machine learning techniques like KNN, ANN, and SVM.

All the results have differences due to the use of different datasets, feature extraction techniques, and machine learning techniques. Most of the datasets used were private and the researchers did not share their data set on the web, or in other words, we can say that those data sets are not publicly available. We have found that Ziaee Hospital recorded Ardakan dataset. The dataset contains 148 PCG signals from 22 subjects (8 males, 14 females; between 3 and 85 years old), unfortunately, the data is not publicly available from the literature.



Fig. 3. Phonocardiography wave signals of arrhythmia-bradyarrhythmia.

TABLE III. DETAILS OF FEATURE EXTRACTION AND ACCURACY

Article	Feature Extraction Technique	Machine Learning Techniques	Dataset	Results Accuracy in %
[32]	MFCC	CNN	PASCAL	87.65
[33]	SEE, HT, MFCC, WT	SVM, KNN	hospital in Ardakan	98.78
[34]	WPT, SVD	CNN, RNN, LSTM	PhysioNet	79.8
[35]	MFCC, Time & Freq	HMM, SVM, CNN	PhysioNet	89.22
[36]	MFCC	DFT, CNN	PhysioNet	86.02
[37]	MFCC	CNN, Springer's algorithm	PASCAL	90.4
[38]	DL	RNN, LSTM, GRU	PASCAL	76.9
[15]	DWT, CWT, STFT, MFCC	SVM, ANN, KNN	PhysioNet	99.7
[39]	MFCC	SVM, RF, MLP	PhysioNet	84.88
[40]	FT, WT, FLP	LVQ, PNN, LS-SVM	PhysioNet	98
[41]	MN, SD, RPAB, SC	SVM	Not available	71.13
[70]	LDA	K-mean	Peter Bentley heart sound	84.39

Another dataset that some researchers have used is GitHub. GitHub dataset they have obtained the PCG signals from a public database but the best result was from GitHub. There are 1,000 PCG recordings present in the database for different subjects. Out of these 1,000 recordings each class contains 200 recordings. There are five classes of PCG signals given in the database namely the healthy control (HC), AS, MS, MR, and mitral valve collapse (MVP). Most of this dataset it's not clear and even the data set is not labeled.

Considered the most popular among the research and the most used in the experiment, the dataset (PASCAL challenge dataset) consists of 312 auscultations collected in the Real Hospital Português (RHP) Maternal and Fetal Cardiology Unit in Recife, Brazil, using the DigiScope. Each inspection is reported for six to ten seconds. The normal count is 200. And the abnormal count of 112. In reference [37] the collected dataset is divided into two parts: Dataset A contains 31 normal heart sounds and 34 abnormal heart sounds. For the training set, 15 normal and 17 abnormal sounds are selected, and the rest is for the test. Dataset B contains 200 normal and 66 murmurs, 100 normal and 33 abnormal for the training, and the rest is left for the test.

After we collected a large number of databases and searched for them, we found that what is publicly available is PASCAL and PhiysoNet. Thus our research based on them in addition to the availability of a new database called (heartbeat sound) through the site "Kaggle" which is the largest site for contests and review of databases.

B. Mel-Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCC) is a feature extraction technique used in speech processing and audio signal analysis. The MFCC algorithm was first proposed by Davis and Mermelstein in 1980 [42]. It has since become one of the most widely used techniques in speech and audio signal processing. The basic intuition behind MFCC is to extract a compact representation of the spectral envelope of an audio signal, which can then be used for further analysis. The spectral envelope is essentially a smoothed version of the power spectrum of the signal, and it contains information about the distribution of energy across different frequency bands [43]. Fig. 4 shows the signal in the Time Domain.

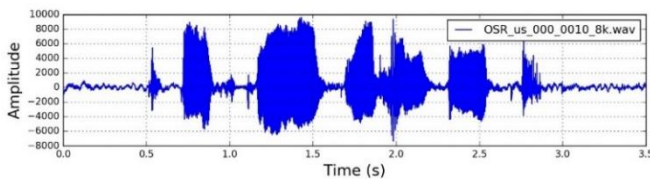


Fig. 4. Signal in the time domain.

To extract the spectral envelope, the MFCC algorithm first divides the signal into short overlapping frames, typically 20-40 MS in duration. Each frame is then transformed into the frequency domain using a Fourier Transform. The resulting frequency spectrum is then passed through a bank of Mel-scale filter banks, which are spaced uniformly on the Mel-scale, a perceptual scale that is more closely related to the way humans perceive pitch than the linear frequency scale [44]. The output of each filter bank is then logarithmically scaled, and the

resulting values are transformed using the Discrete Cosine Transform (DCT) to obtain the MFCCs.

The DCT is used to decorrelate the filter bank outputs and to obtain a set of décor-related coefficients that are more suitable for further analysis [45]. The number of MFCCs extracted depends on the specific application, but typically, 12-13 coefficients are used for speech recognition tasks. The resulting MFCCs can be used as features for machine learning algorithms such as Hidden Markov Models (HMMs), which are commonly used in speech recognition [2].

A Comparative Study on MFCC and MEL spectrogram features for automatic speech recognition presented by J.M. Azevedo, et al. [46]. This study compared the performance of MFCC and Mel spectrogram features for automatic speech recognition using deep learning models. The results showed that Mel spectrogram features outperformed MFCCs in terms of recognition accuracy [46].

Exploring the impact of MFCC and its derivatives on EEG-based emotion recognition was proposed by S. Almogbel, et al. [47]. In this study, the authors investigated the use of MFCC and its derivatives for EEG-based emotion recognition. The results showed that incorporating MFCC and its derivatives improved the accuracy of emotion recognition compared to using raw EEG signals alone [47].

MFCC-based speech enhancement using Deep Neural Networks (DNN) was investigated by Mohamed, et al. [48]. This study proposed a deep neural network-based approach for speech enhancement using MFCC features. The results showed that the proposed approach improved speech quality and intelligibility compared to traditional MFCC-based speech enhancement methods [48]. MFCC and its derivatives-based features extraction for automated speech recognition of spontaneous Tamil language was put forward by S. Sivaprakasam, et al. [49]. This study investigated the use of MFCC and its derivatives for automated speech recognition of the spontaneous Tamil language. The results showed that using higher-order derivatives of MFCCs improved the recognition accuracy compared to using only MFCCs [49].

Comparison of MFCC and Gamma tone filter bank features for speech emotion recognition presented by S. Mohapatra and S. Lenka [50]. This study compared the performance of MFCC and gamma tone filter bank features for speech emotion recognition using machine learning models. The results showed that both features performed similarly in terms of recognition accuracy [50]. An Improved MFCC algorithm for speech recognition lodge by J. Wu, et al. [51]. This study proposed an improved MFCC algorithm for speech recognition by adding a Gaussian filter to the Mel filter bank. The results showed that the proposed algorithm outperformed traditional MFCCs in terms of recognition accuracy [51].

A Comparative Study of MFCC and Gammatone Filter Bank features for phoneme recognition advanced by Y. Sun, et al. [52]. This study compared the performance of MFCC and Gamma tone filter bank features for phoneme recognition using deep learning models. The results showed that Gammatone filter bank features outperformed MFCCs in terms of recognition accuracy [52].

Speaker recognition based on MFCC and XGBoost initiated by Z. Yang, et al. [53]. This study proposed a speaker recognition system based on MFCC features and XGBoost, a gradient-boosting algorithm. The results showed that the proposed system achieved high accuracy in speaker recognition tasks [53]. A Comparative Study of MFCC and DL features for speech emotion recognition argued by X. Liu, et al. [54]. This study compared the performance of MFCC and deep learning features for speech emotion recognition. The results showed that deep learning features outperformed MFCCs in terms of recognition accuracy [54].

MFCC-based speech separation using Convolutional Recurrent Neural Networks tender by M. Hossain, et al. [55]. This study proposed a convolutional recurrent neural network-based approach for speech separation using MFCC features. The results showed that the proposed approach outperformed traditional MFCC-based speech separation methods [55].

C. Machine Learning and Medical Diagnosis

Machine Learning (ML) continues to hold immense significance in medical diagnosis, especially in this digital and AI age. One of its key contributions lies in the increased accuracy it offers. ML algorithms possess the capability to analyze vast amounts of medical data, encompassing patient records, medical images, and genomic information. By identifying complex patterns and relationships within this data, ML models can provide more accurate and reliable diagnoses, often surpassing human capabilities. ML algorithms excel at the early detection and prevention of diseases by analyzing patient data over time, these models can identify subtle indicators and early signs of conditions such as cancer, heart disease, or neurological disorders. Early detection enables healthcare professionals to intervene sooner, improving patient outcomes and potentially saving lives. ML also enables personalized medicine, as it allows for the development of tailored treatment plans based on an individual's unique characteristics, such as genetics, lifestyle, and medical history. By considering these factors, ML algorithms can predict the effectiveness of different treatment options, helping physicians make informed decisions that are specifically catered to each patient's needs [56].

In the field of medical imaging, and medicine phonography ML algorithms have made significant strides. They can

automatically analyze and interpret medical images like X-rays, MRIs, and CT scans, detecting patterns and anomalies that may be challenging for human observers to identify from the image, and sound data. This capability enhances radiologists' efficiency and accuracy, leading to more precise diagnoses and reducing the chances of misinterpretation. ML's integration into electronic health records (EHRs) enables the analysis of large-scale patient data, allowing for data integration and decision support. ML algorithms can uncover hidden correlations between symptoms, risk factors, and treatment outcomes, providing healthcare professionals with valuable insights to make informed decisions. ML-powered decision support systems can suggest diagnoses, treatment plans, and medication recommendations, ultimately improving the overall quality of patient care [57].

Furthermore, ML contributes to increased efficiency and cost savings in healthcare. By automating time-consuming tasks like data entry, documentation, and administrative processes, ML frees healthcare providers to focus more on direct patient care, reducing the burden of paperwork and enhancing operational efficiency. Additionally, ML helps optimize resource allocation and treatment pathways, potentially reducing healthcare costs.

ML process passes through many phases like pre-processing, learning, and evaluation. Data is mostly in the form of unstructured, inconsistent, incomplete, redundant, heterogeneous, as well as noisy. Using techniques like data cleaning, extraction, fusion, transformation, etc. data pre-processing helps prepare raw data usable and consistent for the subsequent phases. The learning phase selects algorithms of learning and refines the parameters of the model to get the required results through the usage of the pre-processed input data. The evaluation phase follows to ensure the attained performance of the learning model, i.e. performance evaluation of the learning algorithm included the selection of dataset, error estimation, measuring performance, and the different tests of statistics. The results of the evaluation phase may lead to the parameters of adjusting for the selected learning algorithm or selecting various algorithms and classifiers. Through the multiple facets like nature of learning, learning target, type of input data, data availability timing, users (stakeholders), and domain factors. Fig. 5, illustrates the Machine Learning framework.

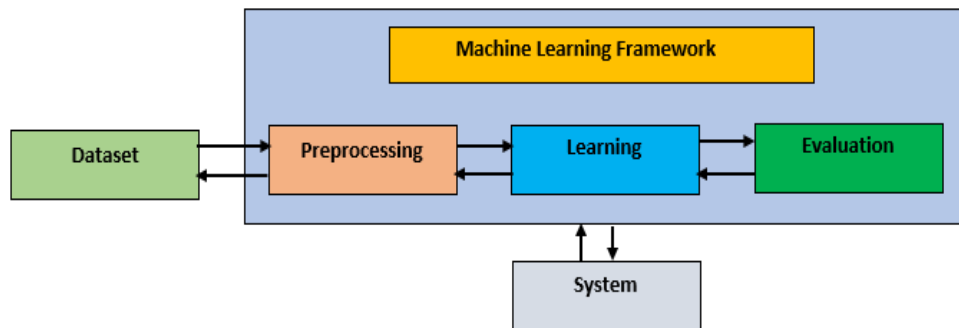


Fig. 5. Machine learning framework (Rizvi, 2019).

D. Supervised Learning

There are three major administrations of conventional machine learning which are supervised, unsupervised, and semi-supervised learning. Supervised learning (SL) is claimed as the uttermost decisive and intrusive annex of machine learning (ML) and pattern recognition. In supervised learning, the learning system is commenced with examples of input and output braces and the prime objective is to learn a function that decorously maps inputs and outputs [58].

Supervised machine learning which is also known as a classification learning advent is employed for analyzing training or labeled data to draft concealed and unseen instances of data for future and imminent classification. To train a classifier that learns to differentiate between different pattern classes, extracted features from recognition units are used [59]. The supervised learning approach carves an acceptable amount of training data or labeled data for the classification of unseen test data [1].

SL, which is also called sometimes a classification learning advent, is implemented for data labeling or training analysis for drafting unseen and concealed data instances for both the imminent and future classification. In training, a classifier for learning to differentiate between the distinct pattern's classes and extracted features from the units of recognition are utilized. The SL approach carves an acceptable amount of training data or labeled data for the classification of unseen test data [59]. The SL model is shown in Fig. 6.

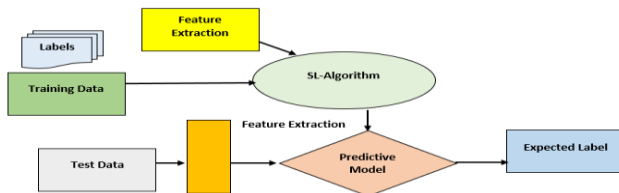


Fig. 6. Supervised machine learning model [59].

E. Artificial Neural Network

Being an SL-based information processing discipline, Artificial Neural Network (ANN) is used for solving circuitous problems. Besides, ANN facilitates understanding the conduct of complex systems through the usage of computer simulations. The final objective of the ANN paradigm is to solve computational problems in the way, the human brain would do. A very simple ANN comprises abounding elementary nodes/processors, often dubbed neurons. These neurons cause a catenation of real-valued activation. Apart from that, sensory input is evoked by the input neurons through the perception of the environment. In the same way, other neurons receive the input of activation from weighted connections of the past antecedent neurons. In this way, only a few neurons, output neurons, for instance, may affect the environment through various prompting actions [60].

The discipline of ANN can be thought of as a set of massively parallel computing systems comprising many interconnected processors. ANN models try to utilize organizational principles like generalization, learning, and computation in a network of weighted graphs. The nodes being used in this network are artificial neurons. Further, the edges

with the weights are joined with the input and the output neurons. In ANN, the process of learning is the determination of weights that force the ANN to show the expected behavior or output. Comprehending the conduct of neurons may entail an array of computational stages and phases in a non-linear manner. Further, each stage converts the network's accumulated activation. Specifically, the techniques of ANN deputize the credit beyond many such stages. Besides, this paradigm is normally adopted due to its vibrant adequacy of the mapping for nonlinear systems [61].

F. Extreme Learning Machines

Both the DL and the classical ML techniques proved a springboard for yet another efficient approach to learning which was introduced during the last decade. This new approach to learning is termed Extreme Learning Machines (ELM). ELM based on the SL approach has been introduced by Huang et al. [62]. Both the bias of ELM and the spawning of input weights in a random fashion are the source of major anomalies between classical DL and ELMs, which are prone to fast learning speed [63]. On the other hand, ELM is a very straightforward, naive, and efficient algorithm, which stands on the principle of Single Layer Feed-Forward Neural Networks (SLF-FNN).

ELM is a special architecture of a Multi-Layer Neural Network, consisting of a Single-Layer Feed-Forward Neural Network (SLFFNN) equipped with hidden neurons, and designated input weights. Further, it has random bias values in the hidden layer, whereas the output is computed by utilizing single multiplication of the weights of the vector matrix. ELM utilizes SLFFNN due to its sufficiency to advance any continuous function and to assert any discontinuous area. If we compare the efficiency of ELM and traditional Back Propagation Neural Networks (BP-NN), the ELM has far better learning time for N sample data. In ELM settings, parameters being used in the hidden layers are absolute i.e., independent of the data. Further, hidden level parameters for the function of activation are haphazardly generated before the perception of training data [64], [65].

G. Deep Extreme Learning Machines

DELM is normally employed for regression and classification purposes in diverse settings since its rate of learning is very rapid. Further, it is very effective as far as the rate of computational convolution is concerned. Classical ANN algorithms require sophisticated measurements and the learning times should be very slow. Further, they can override the learning model given by Khan [66]. Deep Extreme Learning Machines (DELMS) take advantage and benefit from both ELMs and DL techniques. In the phenomenon of DELMs, hidden layers get enhanced in the network structure of ELM and the random initialization process for the weights of input-layer and initial hidden-layer weights along with the initial hidden layer's bias.

Besides, DELMs utilize a new way to calculate the parameters to be used in all the hidden layers except the first using the Least Square Method (LSM) for calculating the output network weight [67]. DELM consists of the architecture with the multi-layer network which has been distributed in two halves. The first half of these two halves learns the original

data in depth to get the most representative novel data through the usage of ELM-AE. As far as the second half is concerned, it calculates the parameters of the network by utilizing the kernel ELM algorithm [68].

There are two main steps of the Deep Extreme Learning Machines DELMs based classification of image-set. In the first step, the global domain-specific DELM model is found by utilizing training images. In the second step, an initialization building of class-specific DELM is carried out through the usage of global representation. It is to be noted that the encoding of both domain-level, as well as class-specific properties of data, is essential in the performance of both steps [69].

IV. DISCUSSION AND CONCLUSIONS

This review paper explored various studies, research papers, and scholarly works related to the integration of phonocardiography and machine learning. This paper provides a detailed introduction that highlights the importance of phonocardiography and machine learning in the field of healthcare, providing background information on phonocardiography by discussing the components of the phonocardiogram and the challenges associated with signal acquisition and processing. Transitioning to machine learning, the algorithms used in analyzing phonocardiograms and the steps involved in developing a machine learning model.

In the view of the research that we have collected that most of the researchers use Mel-frequency cepstral coefficients (MFCC), which is one of the conventional feature extraction techniques. Looking at the machine learning techniques that researchers have tried, we found that two techniques are the most used and the highest performing, which are support vector machine and Artificial neural network (SVM, ANN).

Review outcomes provide a roadmap for future research by pinpointing underexplored areas, automated detection and classification of heart sounds for the early detection of cardiac abnormalities. Studies in arrhythmia detection, heart sound segmentation, and abnormality classification. The paper concludes with a discussion on the limitations and future directions of phonocardiography and machine learning, addressing challenges and suggesting future research possibilities including deep learning models. The presented literature review provides a comprehensive overview of phonocardiography and machine learning, covering fundamental concepts, applications, and potential advancements in the field.

ACKNOWLEDGMENT

I want to express my profound appreciation and gratitude to the supervisor for their unwavering support and encouragement throughout the development of this paper. Their active participation and valuable insights have played a pivotal role in shaping the direction and content of our research. Additionally, I extend heartfelt thanks to the reviewers for their diligent assessment and constructive feedback, which has significantly enhanced the quality and credibility of this work. This paper would not have been possible without the valuable contributions of both the supervisor and the reviewers, and I

am immensely thankful for their involvement in this research journey.

REFERENCES

- [1] Li, J., Guo, Y., Wang, C., and Li, Z. (2020). "A Robust Speech Recognition System Based on MFCC and CNN." 2020 IEEE 3rd International Conference on Information and Computer Technologies (ICICT), pp. 62-66.
- [2] Li, J. P., Haq, A. U., Din, S. U., Khan, J., Khan, A., & Saboor, A. (2020). Heart disease identification method using machine learning classification in e-healthcare. *IEEE Access*, 8, 107562-107582.
- [3] Chakir, F.; Jilbab, A.; Nacir, C.; Hammouch, A. Phonocardiogram signals processing approach for PASCAL Classifying Heart Sounds Challenge. *Signal Image Video Process.* 2018, 12, 1149–1155.
- [4] Jamil, S., & Roy, A. M. (2023). An efficient and robust Phonocardiography (PCG)-based Valvular Heart Diseases (VHD) detection framework using Vision Transformer (ViT). *Computers in Biology and Medicine*, 106734.
- [5] Chen, W., Sun, Q., Chen, X., Xie, G., Wu, H., & Xu, C. (2021). Deep learning methods for heart sound classification: a systematic review. *Entropy*, 23(6), 667.
- [6] Chang, V., Bhavani, V. R., Xu, A. Q., & Hossain, M. A. (2022). An artificial intelligence model for heart disease detection using machine learning algorithms. *Healthcare Analytics*, 2, 100016.
- [7] Gahane, A., & Kotadi, C. (2022, February). An Analytical Review of Heart Failure Detection based on IoT and Machine Learning. In 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS) (pp. 1308-1314). IEEE.
- [8] Kernbach, J. M., & Staartjes, V. E. (2022). Foundations of machine learning-based clinical prediction modeling: Part II—Generalization and overfitting. *Machine Learning in Clinical Neuroscience: Foundations and Applications*, 15-21.
- [9] Yildirim, M. (2022). Diagnosis of Heart Diseases Using Heart Sound Signals with the Developed Interpolation, CNN, and Relief Based Model. *Traitement du Signal*, 39(3).
- [10] Athreya, A. M., Paramesha, K., Avani, H. S., & Madhu, S. (2022). Neural Networks for Detecting Cardiac Arrhythmia from PCG Signals. In *Intelligent Vision in Healthcare* (pp. 103-115). Springer, Singapore.
- [11] Li, S, Li, F, Tang, S, Xiong, W "A Review of Computer-Aided Heart Sound Detection Techniques", *BioMed Res.* pp: 1–10, 2020.
- [12] Raza, A., Mehmood, A., Ullah, S., Ahmad, M., Choi, G. S., & On, B. W. (2019). Heartbeat sound signal classification using deep learning. *Sensors*, 19(21), 4819.
- [13] Deperlioglu, O. (2021). Heart sound classification with signal instant energy and stacked autoencoder network. *Biomedical Signal Processing and Control*, 64, 102211.
- [14] Khan, M. U., Ali, S. Z. E. Z., Ishtiaq, A., Habib, K., Gul, T., & Samer, A. (2021, July). Classification of Multi-Class Cardiovascular Disorders using Ensemble Classifier and Impulsive Domain Analysis. In 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC) (pp. 1-8). IEEE.
- [15] Li, S, Li, F, Tang, S, Xiong, W "A Review of Computer-Aided Heart Sound Detection Techniques", *BioMed Res.* pp: 1–10, 2020.
- [16] Thalmayer, A, Zeising, S, Fischer, G, Kirchner, J. A. "Robust and Real-Time Capable Envelope-Based Algorithm for Heart Sound Classification: Validation under Different Physiological Conditions", *Sensors* 2020.
- [17] Kapen, P.T.; Yousoufa, M.; Kouam, S.U.K.; Foutse, M.; Tchamda, A.R.; Tchuen, G. Phonocardiogram: A robust algorithm for generating synthetic signals and comparison with real life ones. *Biomed. Signal Process. Control* 2020, 60, 101983.
- [18] Giordano, N.; Knaflitz, M. A Novel Method for Measuring the Timing of Heart Sound Components through Digital Phonocardiography. *Sensors* 2019, 19, 1868.
- [19] Wei, W.; Zhan, G.; Wang, X.; Zhang, P.; Yan, Y. A Novel Method for Automatic Heart Murmur Diagnosis Using Phonocardiogram. In *Proceedings of the 2019 International Conference on Artificial*

- Intelligence and Advanced Manufacturing, AIAM, Dublin, Ireland, 16–18 October 2019; Volume 37, pp. 1–6.
- [20] Malarvili, M.; Kamarulafizam, I.; Hussain, S.; Helmi, D. Heart sound segmentation algorithm based on instantaneous energy of electrocardiogram. *Comput. Cardiol.* 2003, 2003, 327–330.
- [21] Chen, T.E.; Yang, S.I.; Ho, L.T.; Tsai, K.H.; Chen, Y.H.; Chang, Y.F.; Wu, C.C. S1 and S2 heart sound recognition using deep neural networks. *IEEE Trans. Biomed. Eng.* 2017, 64, 372–380.
- [22] Liu, Q.; Wu, X.; Ma, X. An automatic segmentation method for heart sounds. *Biomed. Eng. Online* 2018, 17, 22–29.
- [23] Dixon, P., Pavan, A., & Vinodchandran, N. V. (2018). On pseudodeterministic approximation algorithms. In 43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [24] Kamson, A.P.; Sharma, L.; Dandapat, S. Multi-centroid diastolic duration distribution based HSMM for heart sound segmentation. *Biomed. Signal Process. Control.* 2019, 48, 265–272.
- [25] Renna, F.; Oliveira, J.H.; Coimbra, M.T. Deep Convolutional Neural Networks for Heart Sound Segmentation. *IEEE J. Biomed. Health Inform.* 2019, 23, 2435–2445.
- [26] Liu, C.; Springer, D.; Clifford, G.D. Performance of an open-source heart sound segmentation algorithm on eight independent databases. *Physiol. Meas.* 2017, 38, 1730–1745.
- [27] Deng, S. W., & Han, J. Q. (2016). Towards heart sound classification without segmentation via autocorrelation feature and diffusion maps. *Future Generation Computer Systems*, 60, 13–21.
- [28] Aziz, S., Khan, M. U., Alhaisoni, M., Akram, T., & Altaf, M. (2020). Phonocardiogram signal processing for automatic diagnosis of congenital heart disorders through the fusion of temporal and cepstral features. *Sensors*, 20(13), 3790.
- [29] Kumar, V., Arora, A., Suri, P., & Arora, V. (2022). Managing arrhythmias-A guide to Physicians. *J Med Sci Res*, 10(1), 30–38.
- [30] Pugliese, M., Biondi, V., La Maestra, R., & Passantino, A. (2021). Identification and clinical significance of heart murmurs in puppies involved in puppy trade. *Veterinary Sciences*, 8(8), 139.
- [31] Jagannath, D. J., Dolly, D. R. J., & Peter, J. D. (2020). Composite Deep Belief Network approach for enhanced Antepartum foetal electrocardiogram signal. *Cognitive Systems Research*, 59, 198–203.
- [32] Almanifi, O. R. A., Ab Nasir, A. F., Razman, M. A. M., Musa, R. M., & Majeed, A. P. A. (2022). Heartbeat murmurs detection in phonocardiogram recordings via transfer learning. *Alexandria Engineering Journal*, 61(12), 10995–11002.
- [33] Lubaib, P., & Muneer, K. A. (2016). The heart defect analysis based on PCG signals using pattern recognition techniques. *Procedia Technology*, 24, 1024–1031.
- [34] Sujadevi, V. G., Mohan, N., Sachin Kumar, S., Akshay, S., & Soman, K. P. (2019). A hybrid method for fundamental heart sound segmentation using group-sparsity denoising and variational mode decomposition. *Biomedical Engineering Letters*, 9(4), 413–424.
- [35] Noman, F., Ting, C. M., Salleh, S. H., & Ombao, H. (2019, May). Short-segment heart sound classification using an ensemble of deep convolutional neural networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1318–1322). IEEE.
- [36] Krishnan, P. T., Balasubramanian, P., & Umapathy, S. (2020). Automated heart sound classification system from unsegmented phonocardiogram (PCG) using deep neural network. *Physical and Engineering Sciences in Medicine*, 43(2), 505–515.
- [37] El Badlaoui, O., Benba, A., & Hammouch, A. (2020). Novel PCG analysis method for discriminating between abnormal and normal heart sounds. *Irbm*, 41(4), 223–228.
- [38] Li, H., Wang, X., Liu, C., Zeng, Q., Zheng, Y., Chu, X., ... & Karmakar, C. (2020). A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection. *Computers in biology and medicine*, 120, 103733.
- [39] Karan, B., Thakur, G., Rath, A., & Sahu, S. S. (2021, September). Heart Sound Abnormality Detection using Wavelet Packet Features and Machine Learning. In *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)* (pp. 310–314). IEEE.
- [40] Dhar, P., Dutta, S., & Mukherjee, V. (2021). Cross-wavelet assisted convolution neural network (AlexNet) approach for phonocardiogram signals classification. *Biomedical Signal Processing and Control*, 63, 102142.
- [41] Chen, T. E., Yang, S. I., Ho, L. T., Tsai, K. H., Chen, Y. H., Chang, Y. F., ... & Wu, C. C. (2016). S1 and S2 heart sound recognition using deep neural networks. *IEEE Transactions on Biomedical Engineering*, 64(2), 372–380.
- [42] Davis, S.B., and Mermelstein, P. (1980). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences." *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 4, pp. 357–366.
- [43] Ghosh, S., Tripathy, S., Saha, G., and Paul, S. (2021). "An Improved Feature Extraction Method for Speech Recognition using Deep Learning." *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, pp. 728–733.
- [44] Wang, Y., Huang, W., and Zhao, Z. (2020). "MFCC and CQCC-based Automatic Speaker Recognition Using Convolutional Neural Network." *2020 9th International Conference on Systems and Control (ICSC)*, pp. 49–53.
- [45] Nguyen, T.T., Wang, R., and Nguyen, T.T. (2020). "A Comparative Study of Feature Extraction Techniques for Vietnamese Speech Recognition." *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 188–193.
- [46] Azevedo, J. M., Costa, M. J., & Neto, J. P. (2020). A comparative study on MFCC and MEL spectrogram features for automatic speech recognition. *Proceedings of the 2020 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, 407–412.
- [47] Almgöbel, S., Alshammari, R., & Alshamrani, M. (2020). Exploring the impact of MFCC and its derivatives on EEG-based emotion recognition. *Proceedings of the 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, 88–93.
- [48] Mohamed, M. A., Abo-Elsooud, M. A., & El-Said, S. A. (2020). MFCC-based speech enhancement using deep neural networks. *Proceedings of the 2020 International Conference on Computer and Information Sciences (ICCIS)*, 1–6.
- [49] Sivaprakasam, S., Alagappan, K., & Sundararajan, K. (2020). MFCC and its derivatives based features extraction for automated speech recognition of spontaneous Tamil. *Proceedings of the 2020 4th International Conference on Computing Methodologies and Communication (ICCMC)*, 410–416.
- [50] Mohapatra, S., & Lenka, S. K. (2020). Comparison of MFCC and Gammatone filterbank features for speech emotion recognition. *Proceedings of the 2020 International Conference on Advances in Computing, Communication, & Control (ICAC3)*, 1–6.
- [51] Wu, J., Gao, H., & Gao, L. (2021). An improved MFCC algorithm for speech recognition. *Proceedings of the 2021 11th International Conference on Information Science and Technology (ICIST)*, 418–423.
- [52] Sun, Y., Wang, X., & Yu, L. (2021). A comparative study of MFCC and Gammatone filterbank features for phoneme recognition. *Proceedings of the 2021 IEEE 14th International Conference on Advanced Infocomm Technology (ICAIT)*, 178–183.
- [53] Yang, Z., Chen, J., & Yang, Y. (2021). Speaker recognition based on MFCC and XGBoost. *Proceedings of the 2021 IEEE 4th International Conference on Signal Processing Systems (ICSPS)*, 138–142.
- [54] Liu, X., Chen, M., & Fu, Y. (2021). A comparative study of MFCC and deep learning features for speech emotion recognition. *Proceedings of the 2021 IEEE 23rd International Conference on High Performance Computing and Communications (HPCC)*, 367–372.
- [55] Hossain, M., Islam, M. R., & Razzak, M. A. (2021). MFCC-based speech separation using convolutional recurrent neural networks. *Proceedings of the 2021 IEEE 7th International Conference on Computer and Communication Systems (ICCCS)*, 356–361.
- [56] Ahsan, M. M., Luna, S. A., & Siddique, Z. (2022, March). Machine-learning-based disease diagnosis: A comprehensive review. In *Healthcare* (Vol. 10, No. 3, p. 541). MDPI.

- [57] Amjad, M., Raza, H., Muneer, S., & Aslam, M. A. (2022). A Systematic Review on Brain Tumor Detection Using Machine Learning. *Journal of NCBAE*, 1(4), 11-16.
- [58] Masoumian, A., Rashwan, H. A., Cristiano, J., Asif, M. S., & Puig, D. (2022). Monocular depth estimation using deep learning: A review. *Sensors*, 22(14), 5353.
- [59] Rizvi, S. S. R., Khan, M. A., Abbas, S., Asadullah, M., Anwer, N., & Fatima, A. (2022). Deep extreme learning machine-based optical character recognition system for Nastalique Urdu-like script languages. *The Computer Journal*, 65(2), 331-344.
- [60] Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. *Annals of Data Science*, 10(1), 183-208.
- [61] Patel, L., Shukla, T., Huang, X., Ussery, D. W., & Wang, S. (2020). Machine learning methods in drug discovery. *Molecules*, 25(22), 5277.
- [62] Huang, G.-B., Zhu, Q.-Y., Siew, C.K. "Extreme Learning Machine: Theory and Applications", *Neurocomputing*, pp:1-13, 2005.
- [63] Gu, Y., Chen, Y., Liu, J., Jian, X. "Semi-Supervised Deep Extreme Learning Machine for Wi-Fi based Localization", *Neurocomputing*, pp: 282-293, 2015.
- [64] Demertiz, K., Lliadis, L.S, Anezakis, V. "Extreme Deep Learning in Biosecurity: The Case of Machine Hearing for Marine Species Identification", *Journal of Information and Telecommunication*, pp: 492-510, 2018.
- [65] Yao, L., Ge, Z. "Deep Learning of Semi-supervised Process Data with Hierarchical Extreme Learning Machine and Soft Sensor Application", *IEEE Transactions on Industrial Electronics*, pp: 1-8, 2017. DOI: 10.1109/TIE.2017.2733448.
- [66] Khan, M. U., Samer, S., Alshehri, M. D., Baloch, N. K., Khan, H., Hussain, F., ... & Zikria, Y. B. (2022). Artificial neural network-based cardiovascular disease prediction using spectral features. *Computers and Electrical Engineering*, 101, 108094.
- [67] Fayaz, M., Kim, D. "A Prediction Methodology of Energy Consumption Based on Deep Extreme Learning Machine and Comparative Analysis in Residential Buildings" *Electronics*, 1-22, 2017.
- [68] Xiao, D., Li, B., Mao, Y. "A Multiple Hidden Layers Extreme Learning Machine Method and Its Application", *Mathematical Problems in Engineering*, pp. 1-10, 2017.
- [69] Uzair, M., Shafait, F., Ghanem, B., Mian, A. "Representation of Learning with Deep Extreme Learning Machines for Efficient Image Set classification", *Neural Comput & Applic*, pp. 1-13, 2016. DOI: <https://doi.org/10.1007/s00521-016-2758-x>.
- [70] Sujadevi, V. G., Soman, K. P., Vinayakumar, R., & Sankar, A. P. (2017, December). Deep models for phonocardiography (PCG) classification. In 2017 international conference on intelligent communication and computational techniques (ICCT) (pp. 211-216). IEEE.

A Hybrid Classification Approach of Network Attacks using Supervised and Unsupervised Learning

Rahaf Hamoud R. Al-Ruwaili, Osama M. Ouda

Department of Computer Science-College of Computer and Information Sciences,
Jouf University, Al-Jouf, Saudi Arabia

Abstract—The increasing scale and sophistication of network attacks have become a major concern for organizations around the world. As a result, there is an increasing demand for effective and accurate classification of network attacks to enhance cyber security measures. Most existing schemes assume that the available training data is labeled; that is, classification is based on supervised learning. However, this is not always the case since the available real data is expected to be unlabeled. In this paper, this issue is tackled by proposing a hybrid classification approach that combines both supervised and unsupervised learning to build a predictive classification model for classifying network attacks. First, unsupervised learning is used to label the data available in the dataset. Then, different supervised machine learning algorithms are utilized to classify data with the labels obtained from the first step and compare the results with the ground truth labels. Moreover, the issue of the unbalanced dataset is addressed using both over-sampling and under-sampling techniques. Several experiments have been conducted, using the NSL-KDD dataset, to evaluate the efficiency of the proposed hybrid model and the obtained results demonstrate that the accuracy of our proposed model is comparable to supervised classification methods that assume that all data is labeled.

Keywords—Network attacks; supervised learning; unsupervised learning; machine learning

I. INTRODUCTION

The extensive use of the Internet and its continuous development benefit many network users in many aspects. However, in recent years, cyber-attacks have become a growing concern due to their increasing complexity and diversity posing a major threat to governments, businesses, and networks[1]. As a result, network security becomes more important with the widespread use of the network. The purpose of network security is to provide protection and defense against misuse such as modification and unauthorized access.

The task of detecting anomalies in network traffic is experiencing growing demand due to the expanding internet accessibility among individuals[2]. The potential consequences of an intrusion on a computer network encompass a wide range of concerns, including but not limited to the compromise of confidentiality, integrity, and accessibility. These issues can manifest in various ways, such as breaches of privacy or compromise of systems. The primary classifications of intrusion detection systems encompass signature-based detection systems and anomaly-based detection systems[3]. Signature-based systems predominantly depend on established attack signatures to identify and detect unauthorized activities.

In contrast, when encountering unfamiliar attack signatures, the identification of abnormal network activity is mostly conducted through the utilization of anomaly-based technologies.

There are many mechanisms trying to protect the network from outsiders, but these mechanisms can be hacked because the attacker spends enough time and resources to penetrate this perimeter, which may be mostly successful. Despite the multitude of mechanisms implemented to safeguard networks from external threats, determined attackers often invest significant time and resources to breach these defensive perimeters, leading to a high success rate. While firewalls are renowned as one of the most widely used network defense systems, they alone are insufficient in providing comprehensive protection against cyber-attacks. While access control policies play a crucial role in ensuring network security, they can be circumvented through passive authentication methods, thereby undermining their effectiveness. Passive authentication attacks pose a significant challenge to network security because they exploit the trust established within the network. By leveraging legitimate user credentials or session information, attackers can effectively bypass the access control policies implemented by firewalls.

This highlights the need for additional security measures beyond traditional firewall systems. To mitigate the risks associated with passive authentication attacks, organizations should consider implementing supplementary security measures such as strong user authentication protocols, encryption mechanisms, and intrusion detection systems. These layers of defense can help detect and prevent unauthorized access attempts, even if attackers manage to bypass the firewall's access control policies. By adopting a multi-faceted approach to network security, organizations can enhance their resilience against evolving cyber threats and minimize the potential impact of successful attacks[4]. On the other hand, encrypting stored data is a way to achieve data-centric security. However, encrypting stored data is not appropriate for all environments and contexts. Despite the many ways of protection, the attackers always find an entrance to fulfill their desires[5]. Thus, there is a need to identify methods for extracting security information from network data.

Most of the existing methods for classifying network attacks assume that the available datasets are labeled. Hence, they utilize supervised learning techniques to classify network packets. However, in real scenarios, network data are not labeled, and hence supervised learning-based classification methods might not be practically useful. Another issue with existing datasets is that most of the available datasets are

highly unbalanced in the sense that the training samples for some classes are much smaller than the samples available for other types. Ignoring this class imbalance increases the chances that the developed model will learn more about classes with large samples in the data set than about classes with fewer samples. This paper aims to address the class balance problem and use machine learning techniques to identify and extract useful security information from network data to classify different types of network attacks. Network attack classification plays a vital role in detecting and mitigating potential threats to network security. Traditional signature-based methods have limitations in identifying new and sophisticated attacks. Thus, the need for predictive models that can effectively identify attack patterns and classify them with a high degree of accuracy emerges [4] which affects the protection of the network.

Machine learning plays a pivotal role in the development and advancement of several domains within the field of IDS. Statistical methodologies and methods are utilized in order to train the model using a set of training data. When faced with unfamiliar data, the system extracts distinctive and hidden patterns from the dataset in order to provide predictions or classifications, thereby forecasting future trends based on the available data[6]. There are two primary categories of machine learning algorithms: unsupervised learning and supervised learning. In the training phase of Supervised Learning, the model is trained using data that includes both the dependent variable and its corresponding outcome. Conversely, in Unsupervised Learning, the model is trained on unlabeled data, which consists of input data without any associated output information[7].

In this study, an unsupervised pooling technique has been devised, utilizing the K-means method for the purpose of detecting and grouping network intrusion. Next, a supervised learning technique was employed using three distinct algorithms: Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machines (SVM), in order to classify attacks. The NSL-KDD dataset was utilized in this study, employing two distinct methodologies, namely oversampling and under sampling, to ensure the maintenance of data balance.

The main contributions of this work are outlined as follows:

- A data preprocessing strategy is provided as well as data sampling techniques that aim at achieving a more accurate representation of the dataset's features, with the ultimate goal of reducing model bias.
- Introduce and compare the utilization of three models, namely Random Forest (RF), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM), in the context of classifying network intrusion attempts. The classification task is performed using unsupervised learning through the application of the K-means algorithm.
- Presenting a complete review of network intrusion attacks, focusing on the datasets used in previous research. The analysis evaluates the accuracy of four different models, identifying the most realistic model among them.

The subsequent sections of the paper are structured in the following manner: The literature review is presented in Section II. The suggested method is explicated in Section III. Section IV discusses the evaluation metrics while the discussion of the outcomes is presented in Section V. The final Section VI contains future work and the conclusion.

II. RELATED WORKS

In [1], machine learning was used to detect the occurrence of malicious attacks and introduce a feature-based transfer learning framework and transfer learning approach. They also introduced the feature-based learning approach using a linear transformation, called HeTL. A cluster-enhanced transfer learning approach, called CeHTL, has been proposed to make it more potent to detect unknown attacks and evaluation of learning transfer approaches on shared workbooks. The results show that transfer learning methods improve the detection performance of unknown network attacks compared to baselines.

In [5], the authors cover most of the papers that have been released on the attack and defense of membership inference on ML models. They familiarize MIAs with ML models and present current attack methods. They also rated all MIAs papers next to discuss why MIAs work on ML models and summarize the most current evaluation metrics, datasets, and open-source applications of common approaches.

In [4], the researchers proposed a machine learning approach to classify and predict types of DDoS attacks. The authors also used Random Forest and XGBoost classification algorithms. The UNWS-np-15 dataset was extracted from the GitHub repository and Python was used as a simulation. After applying the machine learning models, they generated a confusion matrix to determine the performance of the model. For the Random Forest algorithm, the results show that both Precision (PR) and Recall (RE) are ~89%. For the XGBoost algorithm, the results show that both Precision (PR) and Recall (RE) are about 90%.

The researchers in [8] proposed an efficient framework that learns minimal temporal preferential attack targeting the LSTM model with electronic medical record inputs, they also proposed an efficient and effective framework that identifies sensitive locations in medical records using adversarial attacks on deep predictive models. The results showed weakness in the deep models, as it was more than half of patients can be successfully attacked by changing only 3% of the recording sites with maximum perturbation less than 0.15 and mean perturbation less than 0.02.

In [9], the researchers suggested a stack-based ensemble approach to obtain reliable predictions by combining different algorithms. A powerful processing model called Graphlab Create (GC) was used to perform experiments involving many cases. Recent datasets consisting of attack types were compiled from the UNSW NB-15 and UGR'16 datasets.

In [10], SVM models detect malicious behavior within low-power, low-speed, and short-range networks. They evaluated two SVM approaches, namely C-SVM and OC-SVM. Actual network traffic was used along with the specific network layer attacks that they have implemented to generate and evaluate

VPM detection models. They show that C-SVM achieves a classification accuracy of 100% when evaluated with unknown data taken from the same network topology in which it was trained and an accuracy of 81% when running in unknown topologies.

In [11] the authors proposed the first survey of its kind on adversarial attacks on machine learning in network security. They discussed aggressive attacks against deep learning in computer vision only. They introduced a new classification of adversarial attacks based on machine learning applications in network security and developed a matrix to correlate different types of adversary attacks with a classification-based classification to determine their effectiveness in causing misclassification. A new idea of the adversarial risk network map concept was presented for machine learning in network security.

In [12] they compared different classifiers in the NSL-KDD dataset for binary and multiclass classification. SVM, random forest, and LSTM-RNN model were considered. They show that the proposed model produced the highest accuracy rate of 96.51% and 99.91% for binary classification using 122 features and an optimal set of 99 features, respectively. LSTM-RNN obtained higher accuracy than SVM in binary classification.

In [13], the researchers presented an exploration of how adversarial learning can be used to target supervised models by generating adversarial samples using the Jacobian-based Saliency Map attack and an exploration of classification behaviors. An authentic power system dataset was used to support the experiments presented. The classification performance of two widely used classifiers, Random Forest and J48, decreased by 6 and 11 percentage points when hostile samples were present.

In [14], the authors aimed to detect distributed denial-of-service (DDoS) attacks on financial institutions by using banking datasets. They used multiple classification models to

predict DDOS attacks. Some complexity has been added to the architecture of the generic models to enable them to perform well and application of support vector machine (SVM), k-nearest neighbors (KNN), and random forest (RF) algorithms. SVM showed an accuracy of 99.5%, while KNN and RF recorded an accuracy of 97.5% and 98.74%, respectively, for detecting DDoS attacks. When compared, it is concluded that SVM is more powerful compared to KNN, RF, and existing machine learning (ML) and deep learning (DL) approaches.

In [15], the authors applied the MeanShift algorithm to detect an attack in a network traffic dataset and evaluated the performance of the MeanShift algorithm by two metrics. These metrics are detection rate and detection accuracy. The results of this study showed that the detection rate of the MeanShift algorithm was 79.1 percent, and the detection accuracy of the MeanShift algorithm was 81.2 percent.

In [16], the authors proposed a method for infiltration detection based on deep neural networks. They trained the encoder block based on self-supervised variance learning using unclassified training patterns. Then they inserted the resulting representation into the classification header which was trained using a labeled data set.

In [17] they implemented the machine learning-based detection, classification, and investigation of flood DDoS attacks. They used four supervised learning methods (CART, K-NN, QDA, GNB) and implemented them well, but CART outperforms others based on the investigations that have been conducted.

In [18] they performed a comparative study to analyze the performance of ML algorithms for intrusion detection on the NAL-KDD dataset. They selected only the relevant features. They concluded a reliable identity detection system capable of real-time intrusion detection using different. Table I summarize some related works and provide a comparison between different approaches.

TABLE I. SUMMARY AND COMPARISON OF THE RELATED WORKS

Ref.	Author & year	Study name	Method or Technique	Dataset	Accuracy	Notes
[1]	Zhao,Juan et al. 2019	Transfer Learning for Detecting Unknown Network Attacks	transfer learning approach HeTL and CeHTL	NSL-KDD	0.93%	Assumed The Data Is Labelled.
[4]	Mohmand, Muhammad Ismail et al.2022	A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks	Was Used Random Forest and XGBoost Classification Algorithms	KDD, UNWS-np-15	90%	They Got Good Accuracy Using XGBoost Algorithm But by Using Other Algorithms The Accuracy Was Low
[9]	Rajagopal, Smitha et al.2020	A predictive Model for Network Intrusion Detection Using Stacking Approach	model Graphlab Create (GC) was used	UNSW NB-15 and UGR' 16	95%	It Is Only Limited To Use Of The Graphlab Construct (GC) Model.
[10]	Ioannou, Christiana et al.2021	Network Attack Classification in IoT Using Support Vector Machines	C-SVM and OC-SVM	-	100%	
[12]	Muhuri, Pramita Sree et al.2020	Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks	SVM, random forest, and LSTM-RNN model were considered.	NSL-KDD	99.91 %	LSTM-RNN performs Poorly. In this Experiment The Training Time Was Not Recorded
[14]	Islam, Umar et al.2022	Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models	Has been used multiple classification models to predict DDOS attacks and SVM, KNN, RF algorithms	Bank Dataset	99.5 %	This Model Is Limited To Offline Datasets

[15]	Kumar, Avinash et al.2020	Network Attack Detection Using an Unsupervised Machine Learning Algorithm	MeanShift algorithm	KDD 99	81.2 %	MeanShift Algorithm used Did Not Detect The R2L and U2R Attack Types.
[16]	Lotfi, S et al.2022	Network Intrusion Detection with Limited Labeled Data Using Self-supervision	Supervised and Self-supervised	UNSW-NB15	%94.05	Detect Intrusion With Limited Number Of Labeled Data
[17]	Sangodoyin, Abimbola O. et al.2021	Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning	Supervised Learning Methods and (DA,NB,DT, k-NN)	Dataset For The SDN Classes Of Events (Normal, TCP,HTTP, UDP)	%98	Only DDoS Attack Was Used, This Study Was Limited To The Supervised Learning Method, CART and k-NN Their Hyperparameters Have Not Been Tuned
[18]	Masoodi Faheem et al.2021	Machine Learning for Classification analysis of Intrusion Detection on NSL-KDD Dataset	Comparative Study of Performance Analysis Of Various ML Algorithms	NSL-KDD	%100	U2L Attacks Did Not Produce Enough Results. There Is No Solution To The Problem Of Security Vulnerabilities in Machine Learning Algorithms

III. PROPOSED METHOD

For developing a hybrid classification model for classifying network attacks, a standard approach was followed, relying on the NSL-KDD dataset. Fig. 1 illustrates the workflow that is followed to achieve the goal of detecting intrusions. Initially, the NSL-KDD dataset is acquired, after which data pre-processing takes place. The pre-processing procedure involves many steps that are required to render the data suitable for the algorithms that will be used later. For instance, in this study, data pre-processing includes removing null and duplicated values, in addition to data normalization and fixing the oversampling and under-sampling issues. After that, the data is fed to unsupervised machine learning model, where a K-means algorithm is used, followed by a supervised learning model where three different algorithms are implemented: Random Forest, Support Vector Machine, and K-Nearest Neighbor. Finally, the performance of each of these algorithms is evaluated according to F1 score, involving recall and precision.

A. Dataset

NSL-KDD dataset has been significantly popular in the field of intrusion detection due to its qualities, among many other datasets that are frequently used, whether they are

private, public, or simulated network traffic datasets. Initially, Tavallaee et al. [19] proposed the NSL-KDD dataset as an enhancement of the KDD-99 cup dataset, overcoming its numerous issues, such that the enhanced dataset only contains the selected records from the complete KDD dataset. Despite the enhancement process, the NSL-KDD dataset still suffers from minor issues such as its lack of representation of the low-footprint attacks [20].

The choice fell upon the NSL-KDD dataset since it has less data points than KDD-99, the number of selected records from each difficult level group is inversely proportional to the percentage of records in the original KDD dataset, and it includes no duplicate records in the test set, ultimately leading to better reduction rates. Furthermore, the selected dataset provides less computational expenses for training ML models. Overall, the NSL-KDD dataset contains 41 attributes and one class attribute [21]. The class attribute indicates the type of network traffic, which can be one of five classes: normal, DoS, Probe, R2L, and U2R. The label counts for the NSL-KDD dataset are as follows: normal: 76967, DoS: 52985, Probe: 13954, R2L: 3749, U2R: 252 The dataset has a total of 147,907 rows and 42 columns, with the additional column being the class attribute.

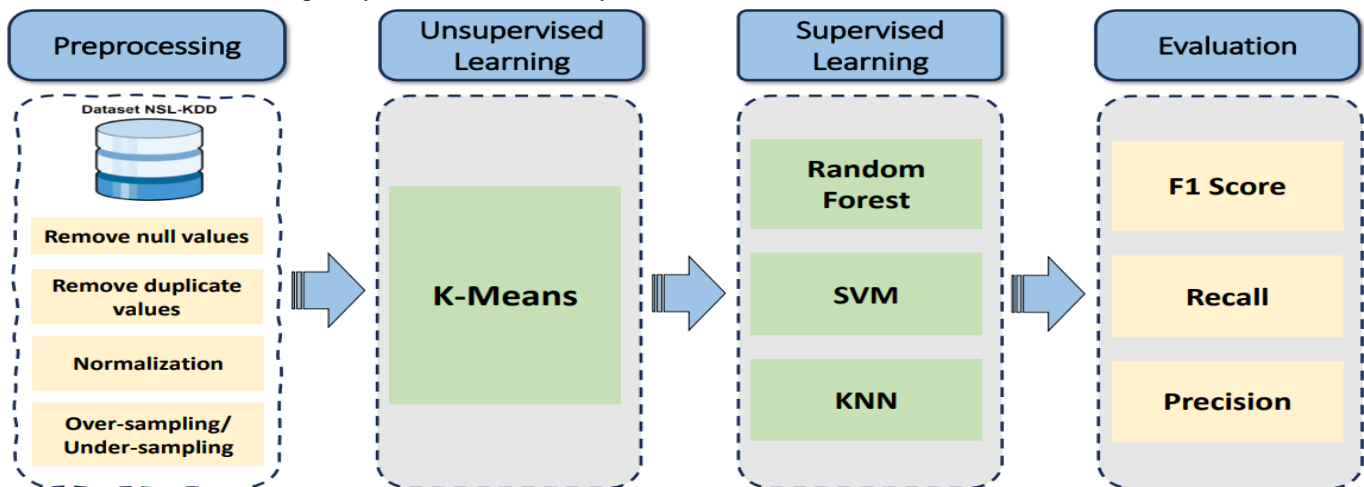


Fig. 1. Proposed framework for the hybrid classification model.

B. Exploratory Data Analysis

Exploratory data analysis EDA is one of the essential steps on any given dataset, as it allows the understanding of the data through observing and analyzing its characteristics, usually by charts. EDA also helps in identifying patterns, possible anomalies, and possible outliers in the data.

The total number of labels within the dataset is 147907 distributed over the following labels: normal, DoS, Probe, R2L, and U2R. The distribution of these labels is presented in Fig. 2 such that the percentage of each label among the whole data is given respectively. Upon observation, the normal label takes up 52% of the total labels which is the highest percentage, followed by 35.8% taken by the DoS label (from the attack labels). Additionally, the Probe labels take up 9.4% of the total labels, while R2L and U2R labels have the lowest percentages.

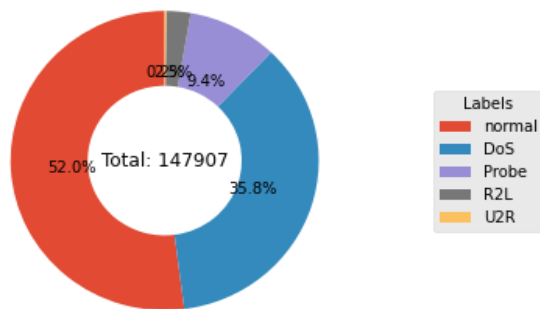


Fig. 2. Labels distribution in the NSL-KDD dataset.

In addition, there are three protocol types within the dataset, namely tcp, udp, and icmp. The percentage of each of these protocol types is shown in Fig. 3. The majority of the protocols are represented by the tcp protocol (82.1%) followed by the udp protocol (11.9%) and the icmp protocol (6.1%).

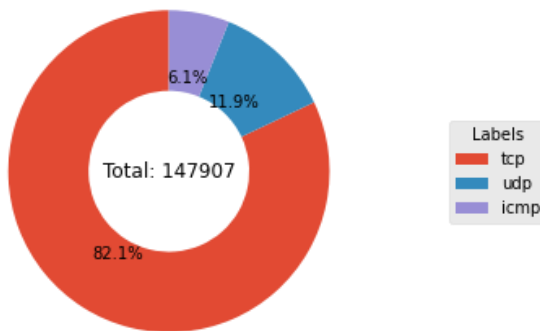


Fig. 3. Protocol type distribution in the NSL-KDD dataset.

It is also possible to know the distribution of the different labels within the dataset over the protocol types that are present. This data is given in Fig. 4. Fig. 4 shows the relationship between the labels and the protocols, where the count plot shows that most of the attacks are carried out using the TCP protocol, with DoS attacks being the most prevalent in this protocol category. It can also be seen that DoS attacks are most prevalent in the UDP protocol, even though UDP doesn't present that many attacks compared to the TCP protocol. Finally, the probe attacks are the most prevalent label within the icmp protocol.

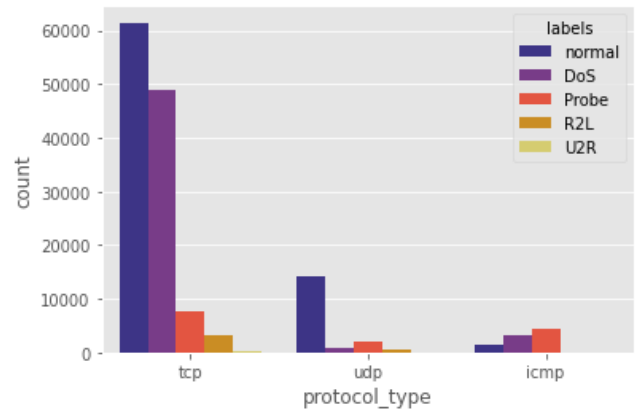


Fig. 4. Label-Protocol distribution count plot.

The relationship between flags and labels can also be acquired from EDA, as presented in Fig. 5. There are numerous flags, namely normal REJ, SF, S0, RSTO, RSTR, RSTOS0, S1, S3, S2, SH, and OTH. It appears that most of the attacks were carried out through the SF flag, where the other dominant flags are REJ flag and S0 flag. The SF flag shows a high count of normal label, whereas the S0 flag shows the highest count of attacks, namely the DoS attacks. It's noteworthy that the DoS attack dominates the REJ flag as well, but in a less prominent way.

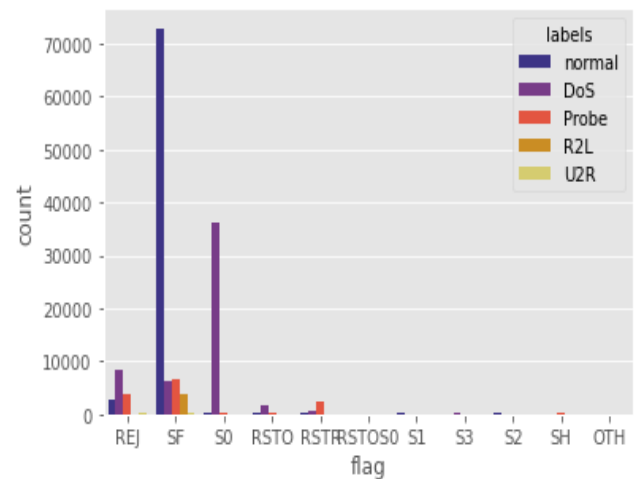


Fig. 5. Label-Flag distribution count plot.

C. Data Pre-processing

Data pre-processing is yet another essential step to prepare the data for use on different machine learning algorithms. The purpose of applying data pre-processing techniques is to improve the quality of data by removing noise and dealing with missing values, for example. It also enhances the efficiency of data analysis and interpretation, while also improving the overall performance of the model. In this study, three main steps were followed as data pre-processing procedures.

1) *Missing values*: The NSL-KDD dataset was checked for missing values or duplicate values. After inspection, it was clear that the dataset does not contain any missing values or duplicate values, which renders it of high quality.

2) *Normalization*: Data normalization works on transforming the data into the same scale without interfering with the relationship between variables or their distribution. This step helps in improving the efficiency and accuracy of the ML model. In this study, the MinMaxScaler function was used to normalize the data, scaling them to a range between 0 and 1.

3) *Class imbalance*: In cases of class imbalance, such as the case of the NSL-KDD dataset, the end results of the ML can be biased. For this reason, the NSL-KDD dataset was subjected to over-sampling and under-sampling to fix the class imbalance issue. Under-sampling is a method used to decrease the number of samples in classes that are overrepresented in a dataset. This can be achieved by randomly selecting a portion of the samples or eliminating samples that have a high degree of similarity to other samples in the dataset. Conversely, oversampling is a technique that aims to increase the number of samples in minority classes by creating synthetic data points.

TABLE II. COUNTS OF EACH CLASS BEFORE AND AFTER SAMPLING METHODS

Sampling Method	Class	Count Before	Count After
Over-sampling	normal	76967	76967
	DoS	52985	76967
	Probe	13954	76967
	R2L	3749	76967
	U2R	252	76967
Under-sampling	normal	76967	252
	DoS	52985	252
	Probe	13954	252
	R2L	3749	252
	U2R	252	252

Table II shows the count of each of the five classes before and after the class imbalance procedures, which are over-sampling and under-sampling. For instance, the normal class contained 76967 instances before under-sampling, and that number became 252 after the procedure was done. Another example is the R2L class which contained 3749 instances before over-sampling, and after the procedure the count became 76967.

The shape of the data before and after the two sampling methods (over- and under-sampling) can be visualized in Table III.

TABLE III. SHAPE OF DATA BEFORE AND AFTER SAMPLING METHODS

Sampling Method	Shape Before	Shape After
Over-sampling	(147907, 122)	(384835, 122)
Under-sampling	(147907, 122)	(1260, 122)

D. Classification Methods

Since there are five different classes or labels within the dataset, this means that the classification problem is a five-class classification problem, where the classes are: benign (or normal), u2r, r2l, dos, and probe. In addition, multiple algorithms exist such that they support this kind of multi-class classification task. Yet, selecting the suitable ML algorithm is the obvious challenge in this case. In this study, initially the cases where labeled data can be used will be considered, which means that supervised machine learning techniques will be used. After that, semi- and un-supervised machine learning techniques will be considered as well.

1) *Supervised classification*: For classifying attacks through supervised classification, three supervised machine learning algorithms were chosen, namely: Random Forest RF, K-Nearest Neighbor KNN, and Support Vector Machine SVM. The dataset is divided by a 80:20 ration for training and testing respectively, upon which these three ML algorithms will be trained and evaluated.

a) *Random forest*: RF is one of the supervised machine learning algorithms, and its concept is randomly creating a forest of decision trees such that the number of the trees directly correlates with the accuracy of performance. Yet, it is noteworthy to consider that creating the forest is different from constructing a decision tree using the information gain or gain index approach. The main difference between Random Forest and Decision Tree algorithms is that in Random Forest, the processes of finding the root node and splitting the feature nodes occur randomly [22]. Two stages are required for RF classification: the forest creation stage where decision trees are created, and the prediction stage where predictions take place.

The Random Forest algorithm creation method involves the following steps:

- (a) Randomly selecting "K" features from the total "m" features, where $k \ll m$.
- (b) Calculating the node "d" among the "K" features using the best split point.
- (c) Splitting the node into daughter nodes using the best split.
- (d) Repeating steps a to c until "l" number of nodes has been reached.
- (e) Building the forest by repeating steps a to d for "n" number of times to create "n" number of trees.

b) *K-Nearest Neighbor*: KNN is described as a non-parametric and lazy learning method. Non-parametric indicates that no assumptions are made about the underlying data distribution, whereas a lazy algorithm indicates the no need for any training data points to achieve model construction. K-nearest neighbors (K-NNs) classifier depends on Manhattan or Euclidean distances to evaluate similarities or differences between instances in the dataset. Often, the Euclidean distance is the metric of choice in KNN classifiers. In KNN, k represents the number of nearest neighbors used

for classification. The algorithm finds the data point with the minimum distance to the test point and assigns it to the same class [23].

Even though KNN is a simple algorithm to implement, it still has the disadvantage of slow prediction time because of calculating the distance between each data point.

The KNN algorithm is implemented by following these steps:

- Loading the data
- Initializing the value of k
- Iterating from 1 to the total number of training data points (*to obtain the predicted class*).
 - Calculating the distance between the test data and each row of training data using a distance metric such as Euclidean, Chebyshev or cosine.
 - Sorting the calculated distances in ascending order based on distance values.
 - Retrieving the top k rows from the sorted array.
 - Determining the most frequent class of these rows
 - Returning to the predicted class.

c) *Support Vector Machine*: SVM is a supervised type of machine learning algorithm in which, given a set of training examples, each marked as belonging to one of the many categories, an SVM training algorithm builds a model that predicts the category of the new example. SVM has the greater ability to generalize the problem, which is the goal in statistical learning. The statistical learning theory provides an outline for studying the problem of gaining knowledge, making predictions, and making decisions from a set of data. SVM is a type of linear and non-linear classifier, which is a mathematical function that can distinguish between two different classes of objects [24]. SVM has the benefit of being capable of managing high-dimensional data and data with non-linear decision boundaries. Nevertheless, its drawback is that it can be computationally demanding and necessitates careful adjustment of the hyperparameters, such as C and the kernel function, to achieve the best possible performance.

Training an SVM algorithm can be achieved with the following pseudocode:

Require: X and y containing the labeled training data, $\alpha \leq 0$ or $\alpha \leq$ partially trained SVM

- $C \leq$ some value (10 for example)
- repeat
- for all $\{x_i, y_i\}, \{x_j, y_j\}$ do
- Optimize α_i and α_j
- end for
- until no changes in α or other resource constraint criteria met

Ensure: Retain only the support vectors ($\alpha_i > 0$)

In SVM, the C value is a regularization parameter that manages the balance between maximizing the margin and minimizing the classification error. The algorithm progressively improves the values of α_i and α_j to locate the support vectors, which are the data points nearest to the decision boundary. After the algorithm concludes, only the support vectors with $\alpha_i > 0$ are maintained.

2) *Unsupervised classification*: The NSL-KDD dataset was also clustered using an unsupervised ML algorithm. K-Means clustering is employed to group similar instances together and new labels are predicted for the instances. The predicted labels will then be used as the target variable, and the instances will be classified using the same supervised ML algorithms (KNN, SVM, and RF). The performance of each algorithm will be evaluated using the same performance metrics utilized in the supervised classification.

a) *K-Means*: K-Means is an unsupervised clustering technique that is frequently employed for partitioning data into k-clusters. The algorithm is iterative and aims to obtain the optimal value for each iteration. Initially, a preferred number of clusters is chosen, and the data points are distributed into k clusters. A greater k produces smaller groups with finer detail, while a lower k results in larger groups with less detail.

The K-Means algorithm can be summarized in two steps that are repeated until the clusters and their means are stable:

- i. Assign each data item to the nearest cluster center. The nearest distance can be calculated using distance algorithms.
- ii. Calculate the mean of the cluster with all data items [23].

IV. EVALUATION METRICS

The performance of the proposed algorithms is evaluated based on their results in the testing dataset. There are several metrics that can be used to evaluate the performance of the models such as precision, recall, and f1 score. Recall is another term used for sensitivity, which resembles the true positive value, which is also the portion of the correctly classified inputs as positive among the entire inputs that should have been classified as positive. Precision is the portion of the true positive classifications over the entirety of the positive results. F-measure is the harmonic mean of the precision and recall and sums up the predictive performance of a model.

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F - Measure = \frac{2 (Precision \times Recall)}{Precision + Recall}$$

True positive is designated by TP. True negative is designated by TN. False positive is designated by FP. False negative is designated by FN.

V. RESULTS

In computer security, intrusion detection encompasses monitoring computer systems and network to look out for any potential threats embodied by malicious activities, security breaches, or unauthorized access. For intrusion detection models, usually NSL-KDD dataset is used since it comprises many network connections that can be classified as normal traffic or attacks of different types. In this study, two different approaches are implemented to detect intrusion, namely the supervised learning approach (through RF, KNN, and SVM), and the unsupervised clustering followed by supervised learning approach. Both approaches are compared according to their performance in terms of accuracy, precision, recall, f1 score, and confusion matrix. By examining the results of these approaches, the strength, and limitations of each one of them becomes clearer, and it can be considered as an insight into building effective intrusion detection systems and identifying areas for future research.

A. Supervised Learning Approach

1) *Random Forest algorithm after over-sampling:* After applying oversampling techniques, the random forest algorithm was able to score a perfect accuracy rate of 100% on the NSL-KDD dataset. Among the five different classes within the dataset (DoS, Probe, R2L, U2R, and normal), the RF model also scored high precision, recall, and f1 score values as can be seen in Table IV. In fact, RF achieves perfect performance when taking into consideration all the evaluation metrics. These results prove that this model accurately detects intrusions and normal connections.

TABLE IV. RANDOM FOREST CLASSIFICATION REPORT / OVER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	100%	1.00	1.00	1.00

Aside from the perfect recall, precision, and f1 score, the RF algorithm also correctly predicted most of the instances except for a few misclassifications that can be seen in the confusion matrix in Table V. The instances in the attacks category (DoS, Probe, R2L, and U2R attacks) were all perfectly classified, yet the misclassifications fall in the normal category. More specifically, 10 normal connections were classified as DoS attacks, 14 as Probe attacks, and 40 as R2L attacks. These misclassifications may be due to the similarities in network traffic patterns between normal connections and certain types of attacks.

TABLE V. RANDOM FOREST CONFUSION MATRIX / OVER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	15602	1	0	0	2
Probe	0	15447	0	0	0
R2L	0	0	15286	0	0
U2R	0	0	0	15400	0
normal	10	14	40	0	15165

2) *K-Nearest Neighbor algorithm after over-sampling:* After over-sampling on the NSL-KDD dataset, the KNN algorithm was able to score a perfect 100% accuracy in classifying intrusions. Similarly, the precision, recall, and F1 score values were very high as can be seen in Table VI. These

results indicate that the KNN algorithm can identify attack classes with ease, while finding some difficulties in correctly identifying all of the normal connections.

TABLE VI. KNN CLASSIFICATION REPORT / OVER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	100%	1.00	1.00	1.00

As for the confusion matrix depicted in Table VII, the results show that the KNN algorithm can correctly identify most of the instances as attacks and normal connections, with a few errors in the normal class. For instance, 21 normal connections were identified as DoS attacks, 46 as Probe attacks, 150 as R2L, and 15 as U2R attacks. This reflects the poor ability of KNN to classify normal connections. On the other hand, KNN was able to correctly identify all of the instances within the DoS, Probe, R2L, and U2R classes.

TABLE VII. KNN CONFUSION MATRIX / OVER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	15596	3	0	0	6
Probe	9	15425	1	0	12
R2L	0	0	15286	0	0
U2R	0	0	0	15400	0
normal	21	46	150	15	14997

3) *Support Vector Machine algorithm after over-sampling:* The achieved accuracy level by the SVM algorithm after over-sampling of the NSL_KDD dataset was 96%. Similarly, all the other metrics reached high values as shown in Table VIII. for all of the attack classes as well as the normal classes. Even in terms of precision, the SVM model achieved lower values scoring 0.97, with 0.96 recall and 0.96 F1-score.

TABLE VIII. SVM CLASSIFICATION REPORT / OVER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	96%	0.97	0.96	0.96

Table IX describing the confusion matrix of SVM algorithm shows that the model achieves acceptable results in the attack classes, where only a few misclassifications took place. on the other hand, it was demonstrated that the model performs poorly in identifying the normal classes, where a lot of misclassifications can be seen. 76 normal connections were mistakenly identified as DoS, 159 were mistakenly identified as Probe, 122 were mistakenly identified as U2R, and 713 were mistakenly identified as R2L. Another class that shows a rather poor performance of SVM is the U2R class, were 1031 instances were mistakenly identified as R2L attacks. The SVM model rather shows a better performance in the other classes.

TABLE IX. SVM CONFUSION MATRIX / OVER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	15511	8	0	0	86
Probe	26	15315	4	17	85
R2L	0	17	14859	149	261
U2R	0	0	1031	14369	0
normal	76	159	713	122	14159

4) *Random forest algorithm after under-sampling*: When under-sampling techniques were followed, the RF model scored 98% accuracy on the NSL-KDD dataset.

Table X illustrates the high values of accuracy, precision (0.98), recall (0.97), and F1 score (0.98) achieved on all the attack classes as well as the normal class. The achieved results, however, were lower than those scored by RF in over-sampling.

TABLE X. RANDOM FOREST CLASSIFICATION REPORT / UNDER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	98%	0.98	0.97	0.98

Interestingly, the confusion matrix of RF after under-sampling, shown in Table XI demonstrates nearly perfect classifications in all classes, including the normal class. There is 1 misclassification only in the DoS class (classified as Probe), and 1 misclassification only in the Probe class, identified as R2L attack. Other than that, the RF forest after under-sampling is so far the only algorithm that perfectly classified all of the normal connections.

TABLE XI. RANDOM FOREST CONFUSION MATRIX / UNDER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	29	1	0	0	0
Probe	0	26	1	0	0
R2L	0	0	17	0	0
U2R	0	0	0	28	0
normal	0	0	0	0	24

5) *K-Nearest Neighbor algorithm after under-sampling*: After performing under-sampling on the dataset, KNN was able to achieve an overall of 96% accuracy in predicting the classes. Tabel XII shows that KNN achieved a good precision (0.97), good recall (0.96) and goof F1-score (0.96).

TABLE XII. KNN CLASSIFICATION REPORT / UNDER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	96%	0.97	0.96	0.96

Furthermore, the confusion matrix for KNN after under-sampling in Table XIII shows that KNN perfectly classified U2R and Normal classes, while it misclassified DoS in 1 occasion (as Probe) only. Probe class was also misclassified only once by KNN as R2L. The most misclassifications achieved by KNN were in the R2L class, where it misclassified 2 of them as DoS.

TABLE XIII. KNN CONFUSION MATRIX / UNDER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	29	1	0	0	0
Probe	0	26	1	0	0
R2L	2	0	15	0	0
U2R	0	0	0	28	0
normal	0	0	0	0	24

6) *Support Vector Machine Algorithm after under-sampling*: After under-sampling, SVM was able to achieve a total of 96% accuracy on the NSL-KDD dataset. Table XIV shows that SVM has good precision (0.97), recall (0.96), and F1 score (0.96).

TABLE XIV. SVM CLASSIFICATION REPORT / UNDER-SAMPLING

Algorithm	Accuracy	Precision	Recall	F1-Score
	96%	0.97	0.96	0.96

In addition, the confusion matrix for SVM shows in (Table XV) that it can perfectly classify normal and U2R labels without any misclassifications. On the other hand, SVM misclassifies DoS in 1 instance as probe, it also misclassifies Probe in 1 instance as R2L. SVM has 2 misclassifications in the R2L label, where 2 instances are falsely classified as DoS.

TABLE XV. SVM CONFUSION MATRIX / UNDER-SAMPLING

	DoS	Probe	R2L	U2R	normal
DoS	29	1	0	0	0
Probe	0	26	1	0	0
R2L	2	0	15	0	0
U2R	0	0	0	28	0
normal	0	0	0	0	24

B. Overall Performance in Supervised Classification

The scores achieved by all of the supervised algorithms “RF, KNN, and SVM” are shown in Fig. 6. The performances are shown in terms of accuracy, Precision, Recall, and F1-score in both cases of over-sampling and under-sampling.

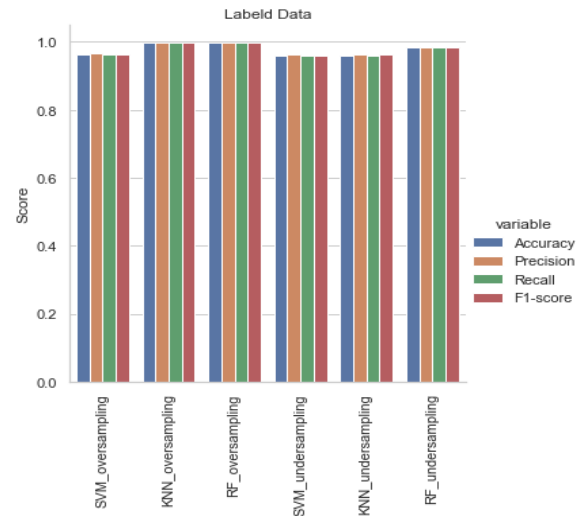


Fig. 6. Performance of supervised ML algorithms in over- and under-sampling.

The results of the three algorithms in terms of accuracy through over-sampling and under-sampling of the NSL-KDD dataset are shown in Table XVI.

TABLE XVI. ACCURACY RESULTS FOR SUPERVISED LEARNING ALGORITHMS IN OVER-SAMPLING AND UNDER-SAMPLING

	KNN	SVM	Random Forest
Over Sampling Accuracy	100%	96%	100%
Under Sampling Accuracy	96%	96%	98%

C. Overall Performance in Unsupervised Classification

The outcomes of the unsupervised categorization employing K-Means indicated that the precision of the supervised algorithms fluctuated based on the sampling method utilized. Following oversampling, the SVM and Random Forest algorithms attained an accuracy of 0.94, whereas KNN attained an accuracy of 0.92. In contrast, following under-sampling, KNN achieved an accuracy of 0.92, while SVM and Random Forest attained an accuracy of 0.94. These results are presented in

XVII.

TABLE XVII. PERFORMANCE OF ALGORITHMS AFTER K-MEANS CLUSTERING AS UNSUPERVISED CLASSIFICATION

	KNN	SVM	Random Forest
Over Sampling	92%	94%	94%
Under Sampling	92%	94%	93%

In addition, the same results can be visualized in Fig. 7 which shows the accuracy, Precision, Recall, and F1-score values for SVM, KNN, and RF after K-means clustering, in both over- and under-sampling techniques on the NSL-KDD dataset.

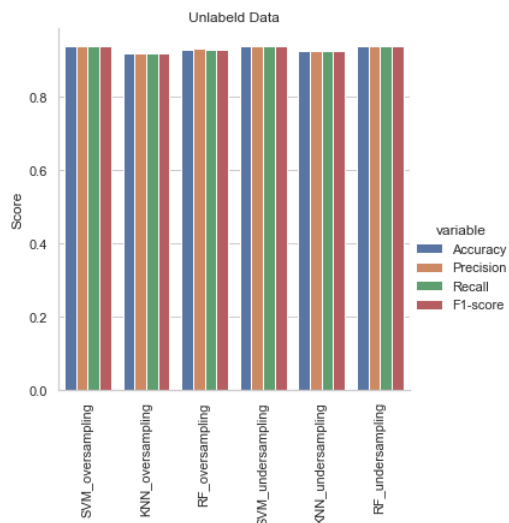


Fig. 7. Performance of unsupervised classification in over- and under-sampling.

VI. CONCLUSION AND FUTURE SCOPE

Intrusion detection is an essential component of cybersecurity and organizations of all sizes to protect their networks and systems from attacks. Effective intrusion detection requires a combination of technical tools and expertise and a thorough understanding of the organization's potential threats and vulnerabilities. This paper proposed a hybrid intrusion detection method that employs both unsupervised and supervised learning to address those issues of unlabeled and unbalanced datasets. Several supervised learning

techniques including Random Forest, K-Nearest Neighbor, and Support Vector Machine were tested along with the K-means unsupervised classification technique. The main task was to perform intrusion detection by classifying traffic data as Normal, DoS, Probe, R2L, and U2R after training the ML algorithms on the NSL-KDD dataset. The obtained results showed that all algorithms can achieve high accuracy, recall, and F1 score. In the future, the integration of ensemble models for classification [25] can be explored. Moreover, the utilization of federated learning to maintain data integrity and privacy [26], and the adoption of transformer ViT models[27] to enhance network attack defense across many datasets can be considered.

REFERENCES

- [1] J. Zhao, S. Shetty, J. W. Pan, C. Kamhoua, and K. Kwiat, "Transfer learning for detecting unknown network attacks," EURASIP J. Inf. Secur., vol. 2019, no. 1, p. 1, Dec. 2019, doi: 10.1186/s13635-019-0084-4.
- [2] A. Devarakonda, N. Sharma, P. Saha, and S. Ramya, "Network intrusion detection: a comparative study of four classifiers using the NSL-KDD and KDD'99 datasets," J. Phys. Conf. Ser., vol. 2161, no. 1, p. 012043, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012043.
- [3] I. P. Saputra, E. Utami, and A. H. Muhammad, "Comparison of Anomaly Based and Signature Based Methods in Detection of Scanning Vulnerability," in 2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Oct. 2022, pp. 221–225. doi: 10.23919/EECSI56542.2022.9946485.
- [4] Y. Wu, D. Wei, and J. Feng, "Network Attacks Detection Methods Based on Deep Learning Techniques: A Survey," Secur. Commun. Networks, vol. 2020, pp. 1–17, Aug. 2020, doi: 10.1155/2020/8872923.
- [5] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang, "Membership Inference Attacks on Machine Learning: A Survey," ACM Comput. Surv., vol. 54, no. 11s, pp. 1–37, Jan. 2022, doi: 10.1145/3523273.
- [6] H. Alazzam, A. Sharieh, and K. E. Sabri, "A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer," Expert Syst. Appl., vol. 148, p. 113249, Jun. 2020, doi: 10.1016/j.eswa.2020.113249.
- [7] et al. Rahim, Rahila, "nalysis of IDS using feature selection approach on NSL-KDD dataset," 2022.
- [8] O. A. Alimi, K. Ouahada, A. M. Abu-Mahfouz, S. Rimer, and K. O. A. Alimi, "A Review of Research Works on Supervised Learning Algorithms for SCADA Intrusion Detection and Classification," Sustainability, vol. 13, no. 17, p. 9597, Aug. 2021, doi: 10.3390/su13179597.
- [9] S. Rajagopal, P. P. Kundapur, and H. Katiganere Siddaramappa, "A predictive model for network intrusion detection using stacking approach," Int. J. Electr. Comput. Eng., vol. 10, no. 3, p. 2734, Jun. 2020, doi: 10.11591/ijece.v10i3.pp2734-2741.
- [10] C. Ioannou and V. Vassiliou, "Network Attack Classification in IoT Using Support Vector Machines," J. Sens. Actuator Networks, vol. 10, no. 3, p. 58, Aug. 2021, doi: 10.3390/jsan10030058.
- [11] et al. Ibitoye, Olakunle, "The Threat of Adversarial Attacks on Machine Learning in Network Security--A Survey," arXiv, vol. arXiv:1911, 2019.
- [12] J. Kumar, R. Goomer, and A. K. Singh, "Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters," Procedia Comput. Sci., vol. 125, pp. 676–682, 2018, doi: 10.1016/j.procs.2017.12.087.
- [13] E. Anthi, L. Williams, M. Rhode, P. Burnap, and A. Wedgbury, "Adversarial attacks on machine learning cybersecurity defences in Industrial Control Systems," J. Inf. Secur. Appl., vol. 58, p. 102717, May 2021, doi: 10.1016/j.jisa.2020.102717.
- [14] U. Islam et al., "Detection of Distributed Denial of Service (DDoS) Attacks in IOT Based Monitoring System of Banking Sector Using Machine Learning Models," Sustainability, vol. 14, no. 14, p. 8374, Jul. 2022, doi: 10.3390/su14148374.

- [15] and R. B. Kumar, Avinash, William Glisson, "Network attack detection using an unsupervised machine learning algorithm," Hawaii Int. Conf. Syst. Sci. 2020, 2020.
- [16] et al. Lotfi, S., "Network intrusion detection with limited labeled data," arXiv, vol. 2209.03147, 2022.
- [17] A. O. Sangodoyin, M. O. Akinsolu, P. Pillai, and V. Grout, "Detection and Classification of DDoS Flooding Attacks on Software-Defined Networks: A Case Study for the Application of Machine Learning," IEEE Access, vol. 9, pp. 122495–122508, 2021, doi: 10.1109/ACCESS.2021.3109490.
- [18] F. Masoodi, "Machine learning for classification analysis of intrusion detection on NSL-KDD dataset," Turkish J. Comput. Math. Educ., vol. 12, no. 10, pp. 2286–2293, 2021, doi: <https://doi.org/10.17762/turcomat.v12i10.4768>.
- [19] M. Tavallae, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Jul. 2009, pp. 1–6. doi: 10.1109/CISDA.2009.5356528.
- [20] J. McHugh, "Testing Intrusion detection systems," ACM Trans. Inf. Syst. Secur., vol. 3, no. 4, pp. 262–294, Nov. 2000, doi: 10.1145/382912.382923.
- [21] B. Ingre and A. Yadav, "Performance analysis of NSL-KDD dataset using ANN," in 2015 International Conference on Signal Processing and Communication Engineering Systems, Jan. 2015, pp. 92–96. doi: 10.1109/SPACES.2015.7058223.
- [22] M. W. Liaw, Andy, "Classification and regression by randomForest," R news 2.3, pp. 8–22, 2002.
- [23] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), May 2019, pp. 1255–1260. doi: 10.1109/ICCS45141.2019.9065747.
- [24] and C.-J. L. Hsu, Chih-Wei, Chih-Chung Chang, "A practical guide to support vector classification," Taipei, Taiwan, pp. 1396–1400, 2003.
- [25] A. M. Al-Hejri, R. M. Al-Tam, M. Fazea, A. H. Sable, S. Lee, and M. A. Al-antari, "ETECADx: Ensemble Self-Attention Transformer Encoder for Breast Cancer Diagnosis Using Full-Field Digital X-ray Breast Images," Diagnostics, vol. 13, no. 1, p. 89, Dec. 2022, doi: 10.3390/diagnostics13010089.
- [26] S. Pandya et al., "Federated learning for smart cities: A comprehensive survey," Sustain. Energy Technol. Assessments, vol. 55, p. 102987, Feb. 2023, doi: 10.1016/j.seta.2022.102987.
- [27] R. M. Al-Tam et al., "A Hybrid Workflow of Residual Convolutional Transformer Encoder for Breast Cancer Classification Using Digital X-ray Mammograms," Biomedicines, vol. 10, no. 11, p. 2971, Nov. 2022, doi: 10.3390/biomedicines10112971.

Violent Physical Behavior Detection using 3D Spatio-Temporal Convolutional Neural Networks

Xiuhong Xu¹, Zhongming Liao^{2*}, Zhaosheng Xu³

College of Photovoltaic Power Generation, Jiangxi New Energy Technology Vocational College, Xinyu 338004 Jiangxi, China¹

Academic Affairs Office, Xinyu College, Xinyu 338004, Jiangxi, China²

School of Mathematics and Computer Science, Xinyu College, Xinyu 338004, Jiangxi, China³

Abstract—The use of surveillance cameras has made it possible to analyze a huge amount of data for automated surveillance. The use of security systems in schools, hotels, hospitals, and other security areas is required to identify the violent activities that can cause social, economic, and environmental damage. Detecting the mobile objects on each frame is a fundamental phase in the analysis of the video trail and the violence recognition. Therefore, a three-step approach is presented in this article. In our method, the separation of the frames containing the motion information and the detection of the violent behavior are applied at two levels of the network. First, the people in the video frames are identified by using a convolutional neural network. In the second step, a sequence of 16 frames containing the identified people is injected into the 3D CNN. Furthermore, we optimize the 3D CNN by using the visual inference and then a neural network optimization tool that transforms the pre-trained model into an average representation. Finally, this method uses the toolbox of OPENVINO to perform the optimization operations to increase the performance. To evaluate the accuracy of our algorithm, two datasets have been analyzed, which are: Violence in Movies and Hockey Fight. The results show that the final accuracy of this analysis is equal to 99.9% and 96% from each dataset.

Keywords—Violence detection; surveillance cameras; 3D Convolutional Neural Network (3D CNN); Spatio-temporal convolution; deep learning; abnormal behavior

I. INTRODUCTION

The perception of human behavior and the analysis of its activities has been faced with many challenges so far [1]; therefore, the processing of movies and also the perception of the movies' content with the proper accuracy and on a large scale has particular importance [2]. In this regard, due to the use of cameras in the city, we are faced with a very large amount of video and its content. It is impossible to analyze this amount of information by humans [3]. By analyzing the recorded frames from the surveillance videos, it is possible to recognize the abnormal behavior and the violent activity that has occurred, and it is possible to make effective decisions at the appropriate time and conditions [4]. In the prior years, with the development in the area of computer vision, a large number of new methods have emerged and have attracted much attention from researchers because of their wide-ranging security applications. In 2017, for example, 954,261 CCTV cameras were fixed at the generic level in South Korea, with an increase equal to 12.9% from the prior year [5]. The target of the installation of the cameras is to provide security in generic

locations. To this end, we concentrate on violence detection with the use of cameras. Violence is an unusual activity that includes the bodily power to harm something or to murder or injure a person or a brute. These operations can be detected by an intelligent monitoring model that can be applied to ban these incidents before they become more deadly. The main application of the security systems, which are fixed in various locations such as schools, hotels, streets, and so on, is to comfort the availability of security guards by alarming them to violent activities. However, the human performance monitor on the surveillance film is too slow, which causes life loss and property. Therefore, there is a request for an automatic violence recognition model [6]. Hence, this area of study is constantly expanding, and different techniques have emerged in this field.

Violence detection has shown its application in the modern systems used by humans due to its wide and significant applications in the field of human security and comfort. By reviewing the existing articles in the field of violence detection, we can refer to the presented method in [7]. This method extracts the proper features by combining the spatiotemporal features and also acceleration features, each of which is obtained by using a two-dimensional convolutional network and then a recursive LSTM network. The acceleration changes are the important components in the detection of person-to-person violence. In this article, this acceleration has been calculated and has been modeled by using the severity of the ocular stream changes in three consecutive frames. To calculate spatial features, this article uses the VGG19 network [8] trained on Imagenet and then selects the features from the penultimate layer as the feature vectors. Another network that considers the changes in the optical flow as a feature is a Tdd network [9] which was trained and created by using the UCF101 dataset.

In another research presented in [10], the idea of TSN, that is, the temporal division network, was introduced. This method was actually a new framework for the detection of movie-based performance, and it was based on the concept of long-range domain structure modeling. Their method was the extraction of the sparse temporal feature, which involves surface video monitoring to enable effective learning by using violent and action-packed videos that sparsely sample the input frames. It can be said that this architecture is segmental. In research [11], the authors used 3D ConvNet and the key-frame to extract the features from clips that contain violent scenes. They have used 3D ConvNet for the short clips, and also, they have used the

key-frame for the longer clips. The key-frame method divides the movie based on the extracted key-frames, and then it examines the similarity between adjacent frames by changing the position of the gray center.

In research [12], the authors present a method for detecting violent robberies from CCTV footage by using a deep end-to-end sequence model. They have used VGG-16 and a pre-trained CNN by the input video frames (which extract features). They process the features trail with two long-term and short-term memories (convLSTM) using LSTM convolutional, which receives the trail of the obtained features. After these steps, in the end, they used several fully-connected layers for the prediction and classification. In this method, the types of firearms and cold weapons in the image are recognized; in this way, the robberies that show different levels of aggression can be classified.

According to the different detection methods that were examined, it was found that there is a gap in the correct extraction of the features; thus, each of the methods has a high computational complexity and a high cost, as well as they have network overhead and the loss of the movement's features. So, to dissolve the moot point of violence detection, the scene information of both levels is needed (namely, the structure of the scene and the movement made by the people present in the scene). Therefore, by examining different methods, a network has been designed, which is fully explained in the following sections. Our contributions include:

- By taking into account the limitations of the presented methods in this field which are presented in the next section, we present a 3D CNN model for learning the complex sequential patterns to accurately predict the violence from video frames.
- A major limitation in the existed methods is the processing of the un-important frames, which leads to the use of the more memory and the very time-consuming. By considering this limitation, we first identified the people in the video stream by using a pre-trained MobileNet CNN model. Only a trail of 16 frames containing the individuals was transferred to the 3D CNN model for the final prediction, which has helped to achieve the efficient processing.
- Next, inspired by the concept of the transfer learning, 3D-CNN was set up by using the standard datasets to detect the violence in the internal and external surveillance.
- After obtaining the trained deep learning model, it was optimized by using the toolbox of OPENVINO to speed up and improve its performance in the phase of the model deployment. By using this strategy, the trained model was transformed into an average representation based on the trained weights and topology.

The research rest is as follows: Section II discovers our method. The empirical evaluation is considered in Section III. The conclusions and suggestions are presented in Section IV.

II. PROPOSED METHOD

In this part, we explain our method. In this method, violent behavior is recognized with the use of the end-to-end deep learning scheme. The general procedure is as follows: The camera records the film sequence. These film frames are sent directly to a trained MobileNet CNN model. This work is done to identify the persons in the film frame. When a person is detected in the film, a sequence of 16 frames is formed, and then it is sent to a 3D CNN to exploit the spatiotemporal features. These extracted spatiotemporal characteristics are fed to the Softmax classifier to examine the extracted characteristic associated with an activity, and then it provides the predictions. When violence is detected in the video frame, an alarm is sent to the nearest police stand. The suggested method is displayed in Fig. 1. In each of the following subsections we illustrate one step of our method.

A. Pre-Processing

To detect the violent behavior, the first step is the step of recognition of the persons in the video frame. Therefore, the first phase in the pre-processing stage is to use the methods to identify the persons. We only process the parts of the video that contain the persons, avoiding the irrelevant frames, instead of processing the total film. The input video is injected into the MobileNet-SSD CNN system [13], and the persons in the video are identified. The presented model uses CNN to identify the persons because it limits the delay and the size. The MobileNet model processes the separable deep convolutions for object recognition. If the deep and point convolutions are numbered separately, then there will be 28 layers, and each layer will have a non-linear Batch of the ReLu type. Of course, the fully-connected layer will not have this feature. The first convolution layer will consist of 2-strides with a filter form equal to $3 \times 3 \times 3 \times 32$ and the size of input equal to $224 \times 224 \times 3$. The subsequent deep convolution has one stride, its filter form is equal to $3 \times 3 \times 32$, and its input size is equal to $112 \times 112 \times 32$. Mainly, the MobileNet model is applied for the classification tasks. Meanwhile, its SSD is applied to put the multi-box recognizer, and it performs a combination of object recognition. This version, for this purpose, is added at the terminal of the network that performs the feed-forward convolution. Also, it generates a constant-size team of marginal boxes to certify the object detection in the video by extracting its feature maps. The boundary box convolution filters are formed using a predicted category and a certain probability for each category. The category which has the highest probability represents the existing object. An example of the detection of the persons existing in a video frame by using the described model is presented in Fig. 2.

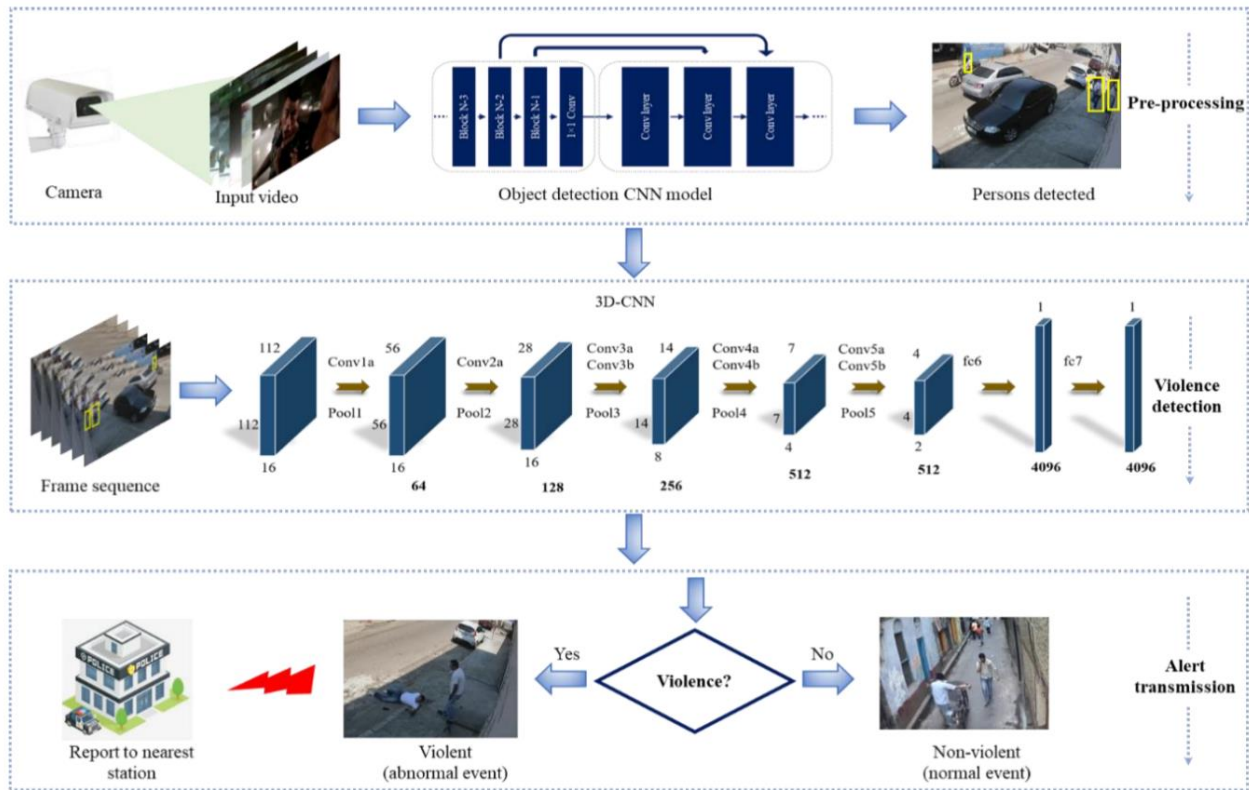


Fig. 1. The presented method for recognition of violent behavior in video frames.



Fig. 2. An instance of the detection of the persons existing in a video frame by using the MobileNet-SSD model.

B. Creation of The 3D CNN Model and Learning Phase

A 3D convolution model is implemented correctly to extract the appropriate spatiotemporal features, and it preserves better temporal information than the Pooling and the 3D convolution operations. There is only spatial information in 2D CNNs, while a 3D CNN can have total temporal information about the input video trail. Some existing methods use the two-dimensional ConvNets. This method is used to exploit the spatial relevance in the film (which simultaneously has temporal relevance). For example, in [14, 15], the 2D CNN model processes several frames. Also, it provides the relevance of total temporal features cumulatively. The 3D convolution model works by convolving a 3D mask in a designed cube (by using assembling the connected frames). In order to capture the motion information, the convolutionally produced feature maps are linked to multiple connected frames on the previous layer. Therefore, the obtained value at the $x.y.z$ location in the map of the feature q on the layer p , which has a bias equal to t_{pq} , is defined by the following relationship:

$$\tanh(t_{pq} + \sum_k \sum_{a=0}^{A_p-1} \sum_{b=0}^{B_p-1} \sum_{c=0}^{C_p-1} \omega_{pqk}^{abc} N_{(p-1)k}^{xyz} = \tag{1}$$

Where C_p is the size of the three-dimensional mask with a time dimension and ω_{pqk}^{abc} is equal to $(a.b.c)$ th value of mask value connected to k th feature mapping on the previous layer. Only the 3D convolutional mask can extract one type of feature because the kernel weights are repeated throughout the cube. Fig. 3 shows the 3D CNN feature maps, which consist of two layers, $conv3a$, and $conv5a$. The presented sample input in Fig. 3 is obtained from the violence class in the dataset.

The average volume of the training data and test data is calculated before starting the training. The proposed network model is well-tuned to take these trails as input. In the Softmax layer, the final divination is calculated as it belongs to the violent category or the non-violent category.

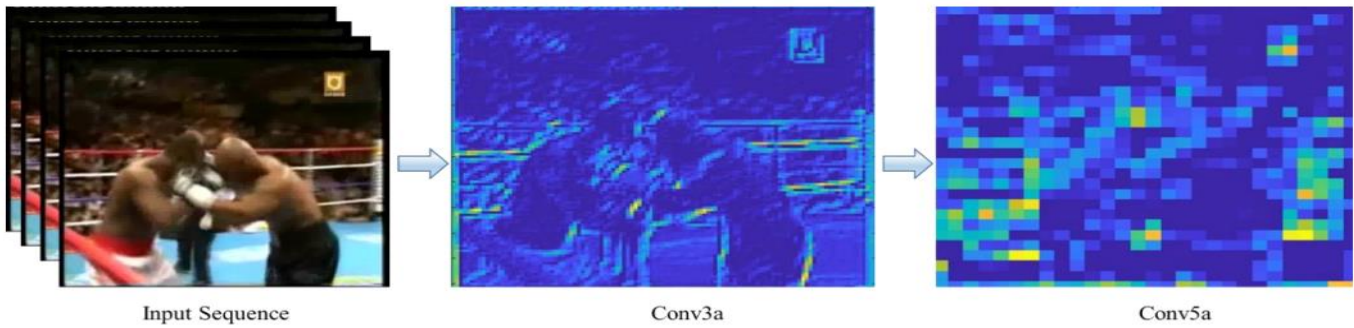


Fig. 3. The 3D CNN feature maps for two layers *conv3a* and *conv5a*.

C. Preparation of Data

To detect the violence, we use a violence dataset that includes a certain number of the stunted film with various times. Each film dataset includes two categories: the violent category and the non-violent category. The entire dataset is separated into a trail of 16 frames with 8-frames overlay among every two consecutive films before starting the training. Then, after obtaining the frames, the total available data is divided into the training set and the test set. A considered ratio for the training set and the test set is equal to 80% and 20%, respectively. When the training set and the test set are taken, a list of files containing paths of the training $L_{tr} = \{S_1, S_{17} \dots S_N\}$ and the test $L_{te} = \{S_1, S_{17} \dots S_N\}$ is generated.

D. Architecture of Proposed Three-Dimensional Convolution

By using the presented 3D CNN model in [16-19], we present our 3D CNN model. The proposed network consists of eight convolutions: five pooling layers, two fully-connected layers, and one Softmax layer. Each convolution layer has a $3 \times 3 \times 3$ kernel by one stride, as well as all pooling layers have a maximum kernel size of $2 \times 2 \times 2$. Of course, the first pooling layer has a kernel size of $2 \times 2 \times 1$ with 2-strides. This exception is to preserve temporal-based data. In the first layer, the number of the considered filters for each convolution, second layer, and third layer, respectively, is 64, 128, and 256. Two fully-connected layers (*fc6*, *fc7*) contain 4096 neurons; also Softmax layer contains N outputs which depend on the number of classes on a dataset. In this paper and in the used dataset, the number of the outputs is equal to two owing to which we have

two categories: the violent scenes and the non-violent. Detailed architecture is shown in Fig. 4.

The proposed architecture catches a trail of 16 frames for input. The input dimensions are equal to 128×171 , but we have used the random cuts with a size equal to $3 \times 16 \times 112 \times 112$ to avoid the problem of overtraining and the problem of not achieving efficient learning. Then, the frames trail is followed by the convolution operations and the 3D pooling operations. The network works as a public feature extractor when the training is done. The various features are trained on the various layer of this network hierarchy. Finally, the exit class is doped as violent or non-violent.

E. Optimization of Proposed Model

The optimization of this model is a process that is applied to produce an optimal design model based on the prioritized limitations. Meanwhile, the power, efficiency, and reliability of this model are maximized. With these methods, we have used the open-source tool of OPENVINO which was created by Intel. This tool develops the workflow through hardware with the maximization of the hardware performance. It runs on the hardware of Intel, and also it takes the prior-trained modules like ONNX, Caffe, TensorFlow, and MXNet as the input, then it transforms them to IR with the use of the model optimizer. Simultaneously and accurately, the model optimizer is applied to make possible a transmission among the training and deployment layers to tune the defined model for optimal implementation in the final model. The stream and the trend of the platform optimization are brief: the model training, the model optimizer, the output (IR), as well as the final platform.

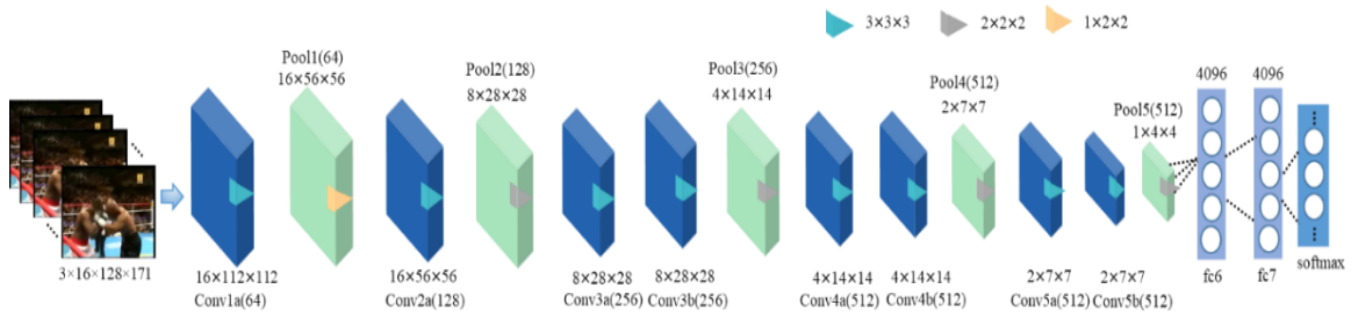


Fig. 4. The structure of the proposed three-dimensional convolution network.

III. TESTS AND REVIEW OF RESULTS

In this part of the article, we present the implementation details of the proposed algorithm, and also we show the performed tests on the dataset and the obtained results. The Python programming language has been used for the implementation of these tests. The presented method is implemented on a computer with a 3.0 GHz Intel(R) Core(TM) i7 CPU and 8G RAM. The convolutional neural network is implemented in GPU, and the graphics card used in this method is NVIDIA GEFORCE 840M.

A. Used Dataset

In this part, the dataset used for the evaluation of our method is explained. In the field of detection of violent behavior, there are different datasets, but in this article, we use two datasets of, Violence in Movies [5] and Hockey Fight [5], which are used in most of the articles in this field.

The dataset of Violence in Movies was nominated by Nievas et al. [5]. This dataset contains 200 videos which include person-to-person combat videos. These scenes are taken from the action movies. Also, in this dataset, the non-combat films are exploited from the datasets of the available act recognition. This dataset discovers the various locations with a mean resolution equal to 360×250 pixels, and each film is restricted to 50 frames. The first person in this dataset, which in the sequence is placed, has little or no camera movement. Similar to the previous dataset, the dataset of Hockey Fight

was nominated by Nievas et al. [5] and consisted of 1000 stunted films obtained from the National Hockey League (NHL). Five hundred clips in this dataset are tagged as violent, as well as 500 clips are tagged as non-violent. Each film contains 50 frames which have a resolution equal to 360×288 pixels. The examples of the frames in these two datasets are mentioned in Fig. 5. The first row is related to the dataset of Violence in Movies, and the second row is related to the dataset of Hockey Fight.

B. Evaluation Results of Proposed Method

In this part, details of the obtained results from the performed experiments on the introduced datasets are presented. Table I displays the results of the performed tests in the dataset of Violence in Movies. As it is known, the highest obtained accuracy is equal to 99.9%, which has a loss equal to 1.67×10^{-7} . This result is obtained at the maximum iterations equal to 5000 with a base learning rate equal to 1×10^{-5} . The important point is that violence detection is easier in the dataset of Violence in Movies than violence detection in the dataset of Hockey Fight. The reason is that there are more people in the clips.

Also, Table II displays the results details of the performed tests on the dataset of Hockey Fight where the highest obtained accuracy is equal to 96% with a loss equal to 5.77×10^{-4} . This result is obtained at the maximum iterations of 5000 as well as the learning rate equal to 0.0001.



Fig. 5. Examples of frames in a used dataset.

TABLE I. THE RESULTS OF OUR METHOD ON THE DATASET OF VIOLENCE IN MOVIES

Learning Rate	Number of Iterations	Loss	Accuracy
0.001	1000	0	99.4%
	3000	0	
	5000	1.21×10^{-2}	
1×10^{-5}	1000	1.99×10^{-3}	99.9%
	3000	5.4×10^{-4}	
	5000	1.67×10^{-7}	

TABLE II. RESULTS OF THE PROPOSED METHOD ON THE DATASET OF HOCKEY FIGHT

Learning Rate	Number of Iterations	Loss	Accuracy
0.001	1000	1.49×10^{-2}	94.9%
	3000	0	
	5000	1.85×10^{-2}	
0.0001	1000	1.79×10^{-3}	96%
	3000	2.27×10^{-3}	
	5000	5.77×10^{-4}	

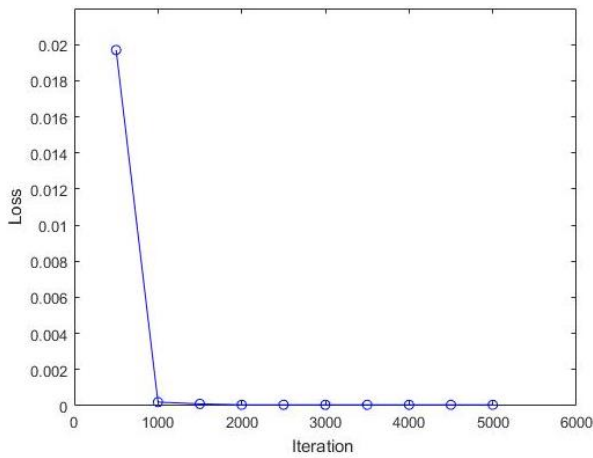


Fig. 6. The trend of the obtained loss value in the dataset of Hockey Fight.

In the experiments, it was found that a learning rate has a significant trace in the amount of the loss. In Fig. 6, the process of changing the loss value is specified. This trend is demonstrated with the change in several learning rates and the iterations equal to 0.001 in the dataset of Hockey Fight. In the 500th iteration of this process, the gained loss value is equal to 1.97×10^{-2} . This value reduces at the same time as the number of iterations. In maximum iteration equal to 5000, this value is equal to 2.32×10^{-7} while the test conditions have not changed, and only the learning rate has changed to 0.0001.

The value of the loss in the dataset of Violence in Movies is very high in the early stages. Then, with increasing in the iteration, the value of the loss is decreased. In this way, the value of the obtained loss in the iteration of 5000 is equal to 5.4×10^{-4} . The trend of the loss reduction in the dataset of Violence in Movies is shown in Fig. 7.

Also, we have appraised the performance of our method by checking the precision, recall, and comparison between datasets with the presentation of AUC in Table III. This table shows the performance of our platform for two datasets. The relationship between the precision and recall is as follows:

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \quad (3)$$

Moreover, the taken confusion matrix is displayed in Table IV. The values of recall as well as precision values for two datasets, respectively, are between $Xmin$, $Ymin$ and $Xmax$, $Ymax$ where X displays precision and Y displays recall for two datasets.

C. Comparison of Presented Method with Similar Methods

One of the important parts of the presentation of a new network is the provision of a complete report of the efficiency of the designed network and the correctness and accuracy of the network performance in different conditions. For this purpose, to demonstrate the better performance and accuracy of the designed network in comparison to similarly designed networks that have been presented and implemented by other researchers, the obtained results from the performance of this network have been compared with other networks. The similarity of the test conditions is related to the same dataset and the same quality evaluation parameters. Therefore, in this part, we contrast the results of two datasets with existing methods. A comparative evaluation with the existing platforms is displayed in Table V. We present the results of the presented method in [20] in the first row. It uses Oriented Violent Flows for motion enlargement. Also, it uses AdaBoost as a feature. Another comparison method that uses Hough forests with two-dimensional CNN for violence detection is presented in [21], and its results are listed in the second row. In addition, we have contrasted the results of our method with the proposed method in [22] that uses motion bubbles and random forests for rapid violence detection. Its results are presented in the third row. Also, in [23], two descriptors are applied in order to identify unusual activities. They applied a simple histogram of the directional tracks together with a dense optical stream in order to detect the abnormal behavior in the terminal result. The fourth row shows the results related to this method. In [24], the writers applied a sliding window method. Also, they use the method of the improved Fisher's vector for violence detection. The results of this method are also displayed in the fifth row.

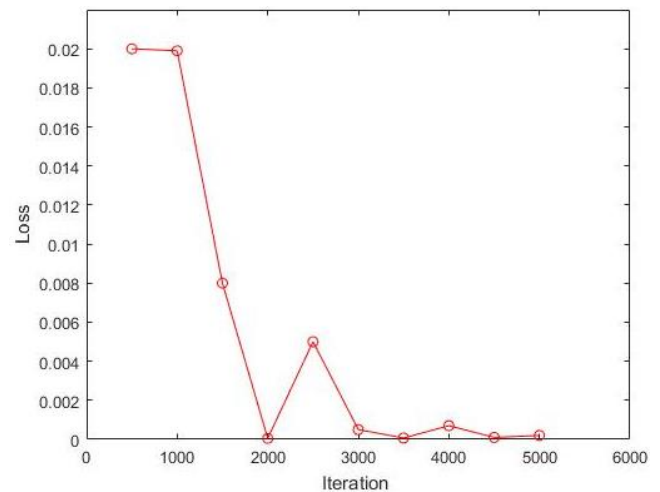


Fig. 7. The trend of the obtained loss value in the dataset of Violence in Movies.

TABLE III. PRECISION, RECALL AND AUC FOR OUR METHOD IN TWO DATASETS

Dataset	Values				Precision	Recall	AUC
	TP	TN	FP	FN			
Hockey Fight	262	230	11	9	0.9957	0.9667	0.97
Violence in Movies	50	57	0	0	1.0	1.0	0.997

TABLE IV. CONFUSION MATRIX

The classes in the dataset	1Hockey Fight		1Violence in Movies	
	Violent	Non-Violent	Violent	Non-Violent
Violent	262.0	11.0	50.0	0
Non-Violent	9.0	230.0	0.0	57

TABLE V. RESULTS OF COMPARISON OF OUR METHOD WITH THE SIMILAR METHODS

Methods	Achieved Accuracies (%)	
	Violence in Movies	Hockey Fight
The method presented in [21]	-	87.5
The method presented in [22]	99	-
The method presented in [23]	96.9	-
The method presented in [24]	98.5	83.1
The method presented in [25]	99.5	93.7
Our proposed method	99.9	96

IV. CONCLUSIONS AND SUGGESTIONS

In this article, the main goal is to identify the violent behaviors in the video frames. Therefore, a three-step method for detecting the violence in the video frames is presented. In this method, first, the people in the video frames are identified. The identification of the people is done by using CNN, which makes the undesirable frames to be ignored and the overhead of the proposed system is reduced. In the second step, a sequence of the frames containing the detected individuals is injected into a trained 3D CNN model, which is performed on two standard datasets. In this dataset, the spatio-temporal features are used and sent to Softmax for the final predictions. Finally, this paper uses the toolbox of OPENVINO to perform the optimization operations to increase the performance. The results of conducted experiments on different datasets show the excellent performance of our proposed platform. These results show that our method is the best suited for detecting the violence in the video surveillance and has better accuracy than several other techniques. For future research, it is suggested that the researchers can ensure that our proposed system can be implemented on the resource-constrained devices. Also, the presented dataset is limited to few violent scenes that can be tried to complete this dataset. On the other hand, some violence is verbal and it is possible to prepare a dataset with this feature in the future works and evaluate the proposed method. In addition, the researchers can propose the edge intelligence to detect the violence in the Internet of Things by using the smart devices for the quick responses.

REFERENCES

- [1] M. S. Ryoo, B. Rothrock, and L. Matthies, "Pooled motion features for first-person videos," 2015, pp. 896-904.
- [2] A. B. Mabrouk and E. Zagrouba, "Abnormal behavior recognition for intelligent video surveillance systems: A review," *Expert Systems with Applications*, vol. 91, pp. 480-491, 2018.
- [3] M. S. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies, "Robot-centric activity prediction from first-person videos: What will they do to me?," 2015, pp. 295-302.
- [4] L. Xia, I. Gori, J. K. Aggarwal, and M. S. Ryoo, "Robot-centric activity recognition from first-person rgb-d videos," 2015: IEEE, pp. 357-364.
- [5] H. H. Park, G. S. Oh, and S. Y. Paek, "Measuring the crime displacement and diffusion of benefit effects of open-street CCTV in South Korea," *International Journal of Law, Crime and Justice*, vol. 40, no. 3, pp. 179-191, 2012.
- [6] E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, and R. Sukthakar, "Violence detection in video using computer vision techniques," 2011: Springer, pp. 332-339.
- [7] W. Yu, K. Yang, Y. Bai, T. Xiao, H. Yao, and Y. Rui, "Visualizing and comparing AlexNet and VGG using deconvolutional layers," 2016.
- [8] C. Cao, Y. Zhang, C. Zhang, and H. Lu, "Action recognition with joints-pooled 3d deep convolutional descriptors," 2016, vol. 1, p. 3.
- [9] C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional lstm network for skeleton-based action recognition," 2019, pp. 1227-1236.
- [10] L. Wang et al., "Temporal segment networks: Towards good practices for deep action recognition," 2016: Springer, pp. 20-36.
- [11] H. Xu, A. Das, and K. Saenko, "R-c3d: Region convolutional 3d network for temporal activity detection," 2017, pp. 5783-5792.
- [12] M. Rezaee, Y. Zhang, R. Mishra, F. Tong, and H. Tong, "Using a vgg-16 network for individual tree species detection with an object-based approach," 2018: IEEE, pp. 1-7.
- [13] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [14] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," *Advances in neural information processing systems*, vol. 27, 2014.
- [15] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthakar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," 2014, pp. 1725-1732.
- [16] Z. Shou, D. Wang, and S.-F. Chang, "Temporal action localization in untrimmed videos via multi-stage cnns," 2016, pp. 1049-1058.
- [17] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," 2015, pp. 4489-4497.
- [18] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Deep end2end voxel2voxel prediction," 2016, pp. 17-24.
- [19] K. Muhammad, T. Hussain, and S. W. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices," *Pattern Recognition Letters*, vol. 130, pp. 370-375, 2020.
- [20] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, "Violence detection using oriented violent flows," *Image and vision computing*, vol. 48, pp. 37-41, 2016.
- [21] I. Serrano, O. Deniz, J. L. Espinosa-Aranda, and G. Bueno, "Fight recognition in video using hough forests and 2D convolutional neural

- network," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 4787-4797, 2018.
- [22] I. Serrano Gracia, O. Deniz Suarez, G. Bueno Garcia, and T.-K. Kim, "Fast fight detection," *PloS one*, vol. 10, no. 4, p. e0120448, 2015.
- [23] H. Rabiee, H. Mousavi, M. Nabi, and M. Ravanbakhsh, "Detection and localization of crowd behavior using a novel tracklet-based model," *International Journal of Machine Learning and Cybernetics*, vol. 9, pp. 1999-2010, 2018.
- [24] P. Bilinski and F. Bremond, "Human violence recognition and detection in surveillance videos," 2016: IEEE, pp. 30-36.
- [25] Bilinski, P.; Bremond, F. Human violence recognition and detection in surveillance videos. In *Proceedings of the 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Colorado Springs, CO, USA, 23–26 August 2016; pp. 30–36.

Construction of VR Video Quality Evaluation Model Based on 3D-CNN

Hongxia Zhao¹, Li Huang²

Information Engineering College, Jiangxi University of Technology, Nanchang, China

Abstract—Currently, virtual reality (VR) panoramic video content occupies a very important position in the content of virtual reality platforms. The level of video quality directly affects the experience of platform users, and there is increasing research on methods for evaluating VR video quality. Therefore, this study establishes a subjective evaluation library for VR video data and uses viewport slicing method to segment VR videos, expanding the sample size. Finally, a classification prediction network structure was constructed using a three-dimensional convolutional neural network (3D-CNN) to achieve objective evaluation of VR videos. However, during the research process, it was found that the increase in its convolutional dimension inevitably leads to a significant increase in the parameter count of the entire neural network, resulting in a surge in algorithm time complexity. In response to this defect, research and design dual 3D convolutional layers and improve 3D-CNN based on residual networks. Based on this research, a virtual reality video quality evaluation model based on improved 3D-CNN was constructed. Through experimental analysis, it can be concluded that the average overall accuracy value of the constructed model is 95.27%, the average accuracy value is 95.94%, and the average Kappa coefficient value is 96.18%. Being able to accurately and effectively evaluate the quality of virtual reality videos and promote the development of the virtual reality field.

Keywords—Virtual reality video; 3D convolutional neural network; residual network; quality evaluation

I. INTRODUCTION

VR panoramic video is a video shot at a full 360° angle with a panoramic camera or cameras. VR panoramic video technology is a real-scene virtual reality technology based on panoramic images, which effectively integrates computer graphics technology, computer simulation technology, sensor technology, display technology and other scientific technologies [1]. VR panoramic video technology covers a variety of content and various forms of video. Among them, the video forms favoured by modern youths, including movies and recorded short films, etc., can be completed by using VR panoramic video technology [2]. At first, VR panoramic video technology was only used for leisure and entertainment, and it was rarely used in other fields. With the continuous development of technology, VR panoramic video technology has entered the fields of education, medicine, tourism, etc. [3]. The video information in mainstream VR is spherical video information, and spherical video has higher requirements for clarity, presence, vertigo, and immersion due to user experience, so timely evaluate VR video quality and improve video design become particularly important [4]. The video quality evaluation can intuitively indicate the quality of the video. 3D convolutional neural network (3D-CNN) is based

on a two-dimensional convolutional neural network, adding a time dimension to the input of the neural network, extracting time and space features at the same time, performing deep learning for behaviour recognition, recognition processing and other operations method, which is widely used in video processing [5]. The research uses the viewport-cutting method to expand the number of videos. Finally, the data set is put into the training of the video quality evaluation model based on 3D-CNN. During the research process, it is found that the increase in the convolution dimension will lead to a sharp increase in the time complexity of the algorithm. In response to this defect, the study designs a double 3D convolutional layer and improves the 3D-CNN based on the shortcut of ResNet. Therefore, a VR video quality evaluation model based on improved 3D-CNN is constructed. Realize accurate and efficient objective evaluation of VR video.

This paper is divided into three parts. The first part is literature review, which analyzes the current research situation at home and abroad in the related research fields involved in the research, summarizes the existing research deficiencies, and points out the future research directions. The second part is the research method part, which expounds the VR video quality evaluation model constructed by the research institute and improves the technology used. The third part is the performance analysis part, which carries out a series of experiments to verify the performance of the model.

The importance and innovation of the research are as follows:

Traditional VR video quality evaluation methods are mainly divided into subjective and objective evaluation. The factors that affect users' perception of VR video quality are not only the perceived quality of the video, but also the subjective feelings such as presence, vertigo and accessibility. However, at present, there are few relevant subjective evaluation databases, and the traditional subjective evaluation is time-consuming and labor-intensive. However, the objective evaluation method needs to compound the subjective evaluation scores and has strong usability. The existing objective evaluation method is more complicated in calculation and its accuracy is not ideal. Therefore, a VR video quality evaluation method based on 3D-CNN was constructed. The research has two innovations, one is to establish a subjective evaluation database of VR video, which includes the quality of perception, the sense of presence, the sense of glare and the acceptability. Two: The existing VR video quality evaluation methods are full reference or partial reference. Firstly, the 3D convolutional channel network is used to evaluate the quality of VR video. Firstly, a visual

interface cutting method is proposed, and a non-reference VR visual frequency quality evaluation method is established by combining 3D convolutional channel network. This method does not need to participate in video, is driven by pure data, does not use human features extraction, and the obtained prediction results are in high consistency with the quality evaluation of V-R video, and the prediction results are good. Compared with the existing full-reference VR video quality evaluation methods, it has stronger competitiveness.

II. RELATED WORKS

With the gradual rise of VR videos, it is becoming more and more important to make accurate judgments on the quality of VR videos. Many scholars have conducted research on video quality evaluation. Tu et al. take user-generated content (UGG) videos as the research object, and comprehensively evaluate the leading no-reference/blind VQA (BVQA) features and models on a fixed evaluation framework. By employing a feature selection strategy on the BVQA model, a new fusion-based quality estimator for AIDeo (VIDEVAL) [6] was created. In order to compare the 8K high-resolution image quality of Versatile Video Coding (VVC) and High Efficiency Video Coding (HEVC) standards, Bonnineau et al. used PSNR, NS-SSIM and VMAF metrics for objective measurement, and obtained the comparative quality evaluation results of the two [7]. Zhang et al. aimed at the problem that many current predictive video quality of experience (QoE) models are too dependent on features specific to a specific feature set and lack generalization capabilities. Using word embedding and 3D convolutional neural networks to extract generalized features and learn them in neural networks, a new end-to-end framework (DeepQoE) was developed to perform classification and regression problems on different multimedia data [8]. Tian et al. used radial symmetric transformations on the luminance components of reference and distorted LF images to explore the depth features of geometric information in LF images. Symmetry and depth features are compared for similarity measurement to obtain video quality scores. It proposed a new full-reference image quality assessment (IQA) method [9]. Lee et al. extracted 3D shear transformation-based spatio-temporal features from overlapping video blocks and applied them to logistic regression, connected with conditional video-based deep residual neural networks to learn spatio-temporal correlations and predict quality scores. It proposed a new frequency-free reference quality assessment method [10]. Yang Aiming at the limitation of the traditional VQA method in capturing complex global time information in a panoramic video, combined spherical convolutional neural network (CNN) and non-local neural network, proposed an end-to-end neural network model for panoramic video and Stereo panoramic video for quality assessment [11].

3D-CNN is a deep learning method based on two-dimensional convolutional neural network, adding a time dimension to the input of neural network, extracting time and space features at the same time, and performing operations such as behaviour recognition and recognition processing. Mzoughi et al. proposed an efficient and fully automatic deep multi-scale three-dimensional convolutional neural network (3D-CNN) architecture based on 3D convolutional layers and deep networks to classify glioma brain tumors [12]. Ramzan et

al. proposed a one-line network for segmenting multiple brain regions based on 3D-CNN, using residual learning and dilated convolution operations to learn an end-to-end mapping from MRI volumes to voxel-level brain segments [13]. The 3D convolutional neural network designed by Liu et al. for the development of Deep-Fake detection has large parameters, resulting in serious memory and storage consumption problems. A lightweight 3D-CNN [14] for DeepFake detection is proposed by using the channel transformation module to extract parameters with fewer features and fusing the spatial features on the temporal dimension for the spatio-temporal module. Hassan-Harrirou et al. In order to reduce the time and cost of exploring the chemical search space in the development of new drugs, more quickly and accurately predict the binding affinity of the lead. An ensemble of 3D-CNNs (RosENet) was used to predict the absolute binding affinities of protein-ligand complexes [15]. Aldoj et al constructed a new semi-automatic prostate cancer classification model using 3D convolutional neural network and histological correlation analysis based on multiparameter magnetic resonance (MR) imaging [16]. Salama et al. took human emotion recognition as the research goal, used the 3D-CNN deep learning framework to extract spatiotemporal features from electroencephalogram (EEG) signals and face video data, and obtained fusion predictions using data enhancement and ensemble learning techniques. The study proposed a new framework for multimodal human emotion recognition [17].

According to the comprehensive literature, there are many video quality evaluation methods, among which 3D-CNN is widely used, but it is not used in 3D video quality evaluation. Therefore, the research builds a VR video quality evaluation model based on 3D-CNN. Evaluate video quality objectively and thoughtfully.

III. CONSTRUCTION OF VR VIDEO QUALITY EVALUATION MODEL BASED ON 3D-CNN

A. Evaluation Index Selection and Video Cutting

VR video is different from traditional video that only records image information from a specific angle per frame. Panoramic video captures image information from all directions at the same time. Through a professional VR head-mounted display, the video is mapped to a spherical surface for users to observe and obtain an immersive experience [18-19]. The VR video transmission framework is shown in Fig. 1.

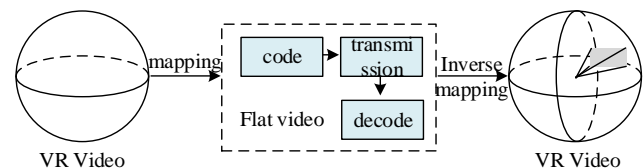


Fig. 1. VR video transmission framework.

In order to analyze video information more efficiently, the research establishes a VR video quality evaluation model based on 3D-CNN based on the characteristics of the human visual system, gives quantitative indicators to analyze video information, and obtains corresponding scores to simulate the

results of subjective evaluation and scoring. Due to the spherical 360° video of VR video, the planar (peak signal-to-noise ratio) PSNR method is usually not accurate enough, so some objective evaluation methods for VR video quality are needed. The study uses Spearman's rank correlation coefficient (SRCC) and Pearson correlation coefficient (PCC) as the correlation evaluation index of subjective and objective evaluation of video quality. The calculation formula of SRCC is shown in the following (1):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

In (1), d_i represents the first i difference, which n is the number of data. The PCC calculation formula is shown in the following (2).

$$\rho(x, y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \quad (2)$$

In (2), $E(X)$ is X the mean value of the variable, and is the mean value $E(Y)$ of the Y variable. The planar video quality evaluation method is not comprehensive. In addition, the study uses spherical-based CP-PSNR and WS-PSNR to obtain more objective evaluation indicators. The weight calculation formula of ERP mapping is shown in the following (3).

$$W(i, j) = \frac{w(i, j)}{\sum_{i=0}^{W-1} \sum_{j=0}^{H-1} w(i, j)} \quad (3)$$

In (3), W and H are the length and width of the video resolution, respectively, and are the scaling factors $w(i, j)$ for pixels to (i, j) be mapped from a plane to a sphere using ERP mapping. The formula is as follows: (4).

$$w(i, j) = \cos\left(\left(j - \frac{H}{2} + \frac{1}{2}\right) \cdot \frac{\pi}{H}\right) \quad (4)$$

The calculation formula of WS-PSNR is shown in (5) below.

$$\left\{ \begin{aligned} WMSE &= \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} (y(i, j) - y'(i, j))^2 \cdot W(i, j) \\ WS - PSNR &= 10 \log\left(\frac{MAX^2}{WMSE}\right) \end{aligned} \right. \quad (5)$$

In (5), $y(i, j)$ and $y'(i, j)$ are respectively the original pixel and the reconstructed pixel, which MAX is the maximum value of the color of the image point. The calculation formula of CP-PSNR is shown in the following (6).

$$\left\{ \begin{aligned} \bar{w}'(s, t) &= \frac{w'(s, t)}{\sum_{s, t} w'(s, t)} \\ CP - PSNR &= 10 \log \frac{I_{\max}^2}{\sum_{s, t} (I(s, t) - I'(s, t))^2 \cdot \bar{w}'(s, t)} \end{aligned} \right. \quad (6)$$

In (6), $I(s, t)$ and represent $I'(s, t)$ the intensity of I_{\max} the point in the reference video and the damaged video respectively, which (s, t) is the maximum value of the color of the image point. In addition to the above-mentioned quality evaluation standards used in general videos, the study establishes a subjective evaluation library for VR videos, and scores the sense of presence, vertigo, and acceptability. A total of 48 VR videos were established in the research database, including 12 source reference videos. Each reference video generated 36 damaged videos through three kinds of QP, and 40 subjects were selected to participate in the establishment of the subjective evaluation database. After the subjects are trained, they evaluate the corresponding VR videos with scores. The evaluation indicators include perceptual quality, sense of presence, vertigo and acceptability. The scores given by the subjects are collected for the evaluation obtained by establishing an objective evaluation model in the future. The score (MOS) is compared with the objective evaluation of the subjects, and a corresponding scoring table is established for subsequent training. Due to the large amount of data required for deep learning, it is necessary to expand the video database, research on cutting VR videos, and cut them into small pieces of video. VR video needs to convert between spherical model and planar model. Before VR video transmission, the spherical model is mapped to the planar model. When the user watches the VR video through the HMD, the planar model is re-projected into the spherical model, so the user actually sees these are the viewpoints of the VR video. After VR video is projected, oversampling will occur. Based on this difference, the study uses the unique viewport characteristics of VR videos to propose a viewport cutting method to cut VR videos. The viewport segmentation diagram is shown in Fig. 2.

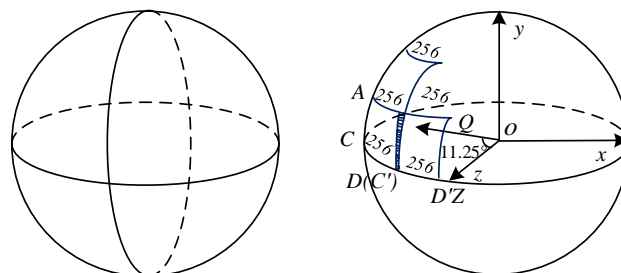


Fig. 2. Schematic diagram of viewport cutting.

Assuming that the user's viewing direction is the negative axis of the X axis and set as the initial position of the head, set the rotation matrix representing the user's rotation relative to the initial position, R and transform the three-dimensional and two-dimensional homogeneous coordinates through intrinsic matrix modeling, such as (7) as shown.

$$K = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \quad (7)$$

In (7), C_x and C_y represent the center point of the viewport texture coordinates, f_x and f_y is the focal length expressed in pixels. The projection relationship formula is shown in the following (8).

$$w \cdot e' = K \cdot R^T \cdot E \quad (8)$$

In (8), E is a point on the spherical surface in the current visible area, e' indicating the two-dimensional homogeneous coordinates of the point mapped on the viewport, and w is the scale factor. By projecting the VR video onto a spherical surface, a series of viewports of VR videos are extracted, and the extracted viewport videos are used as the input of the VR video quality evaluation network.

B. Evaluation Model Construction based on Improved 3D-CNN

After selecting the evaluation index and establishing the evaluation library, the study uses 3D-CNN to evaluate the quality of VR video. The computing modules of traditional convolutional neural networks are mainly divided into four types: convolution, pooling, full connection and classifier prediction. The spatial structure features are extracted through the convolutional layer, and the larger-scale feature learning results are formed through the pooling layer, and then the depth of the convolutional layer and the pooling layer is continuously deepened to extract the spatial features of the image [20-21]. Convolutional neural network (CNN) was originally designed for the feature extraction of two-dimensional data, which can directly establish the mapping relationship from low-level features to high-level semantic features, and has achieved remarkable results in the field of two-dimensional image classification. However, 2D-CNN only carries out sliding calculation on the two-dimensional plane and cannot carry out feature extraction on the spectral dimension of hyperspectral image, so the extracted information is insufficient. A large number of theories and experiments show that 3D-CNN can extract features from both spatial and spectral information dimensions of hyperspectral images to improve the classification

$$a = [M_{1,1,1}, M_{1,1,2}, \dots, M_{1,1,n}, \dots, M_{1,2,1}, M_{1,2,2}, \dots, M_{1,2,n}, \dots, M_{1,m,n}, M_{2,m,n}, \dots, M_{j,m,n}]^T \quad (11)$$

In (11), the original feature map is M_1, M_2, \dots, M_n . The study uses the Softmax regression model at the output layer to realize the multi-category prediction function, and uses the hypothesis function to estimate the probability value of a certain class in the data set. The formula is shown in the following (12).

$$P(y^i = j | x^i; \theta) = \frac{e^{\theta_j^i x^i}}{\sum_{t=1}^D e^{\theta_t^i x^i}} \quad (12)$$

performance of the network. 3D-CNN is a convolutional network with three dimensions, namely image width, image height and image channel. The convolution kernel can move in three directions, and it can be used to better capture the temporal and spatial feature information in the video. The convolution operation formula is shown in the following (9).

$$f(x, y) * w(x, y) = \sum_{s=-a}^a \sum_{t=-b}^b w(s, t) f(x-s, y-t) \quad (9)$$

In (9), it represents $f(x, y)$ the gray value of the point whose $w(x, y)$ coordinates are in the image, which (x, y) is the convolution kernel. Sliding on the image through the weight window, the weighted sum of the pixels on the image and the weight window is used to extract the advanced features of the image. The number of input and output feature maps of the pooling layer is the same, and the calculation matrix form of the pooling layer is expressed as (10).

$$vec(y) = S(x)vec(x) \quad (10)$$

In (10), $vec()$ represents the vectorization operation, and $S(x)$ the feature selector matrix of the input feature map for the pooling layer. In order to prevent the fitting phenomenon, reduce the number of parameters, and perform maximum pooling on the image, the process is shown in Fig. 3.

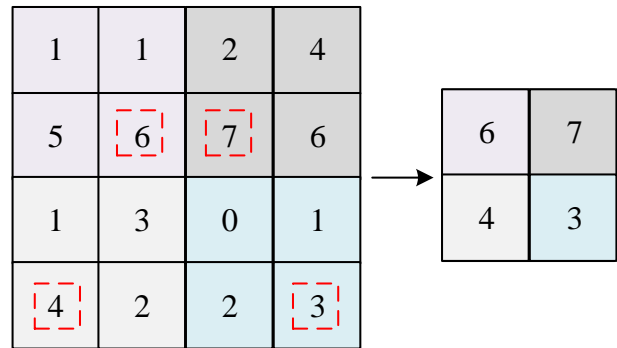


Fig. 3. Maximum pooling process.

After the pooling layer, since the output is a two-dimensional feature map, the fully connected layer vectorizes the map, and the obtained vector is as in (11).

In (12), θ is the model parameter, D is the number of categories, y is the label, and x is the test input. The cost function of Softmax is as follows (13).

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{k=1}^D 1\{y^i = k\} \log \left(\frac{e^{\theta_k^i x^i}}{\sum_{t=1}^D e^{\theta_t^i x^i}} \right) \right] + \frac{\lambda}{2} \sum_{i=1}^m \sum_{j=0}^n \theta_{ij}^2 \quad (13)$$

In (13), $1\{\}$ is the indicative function, where $1\{true\} = 1, 1\{false\} = 0$ is the coefficient of the penalty term.

λ The partial derivative of the cost function to the parameters is as follows (14).

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m [x^i 1\{y^i = j\} - P(y^i = j | x^i; \theta)] + \lambda \theta_j \quad (14)$$

The optimal solution for global convergence is obtained by minimizing the gradient descent algorithm. $J(\theta)$ Traditional convolutional neural networks can only extract spatial features in images, and are not suitable for video processing. 3D convolutional neural networks extract temporal features in videos based on traditional convolutional neural networks. And use the spatio-temporal features to classify and then predict. First of all, study the use of 3D-CNN to form a ten-category network structure. First, divide VR videos into ten categories through MOS points, of which 1-10 points are divided into one category, 10-20 points are divided into two categories, and so on, 90- 100 points are divided into 10 categories. 3D-CNN extracts temporal and spatial features based on 2D convolutional neural networks. The classification structure of 3D-CNN consists of eight 3D convolutional layers,

five 3D maximum pooling layers, two fully connected layers and a ten-category output layer. After the output layer is calculated by softmax, the classification result is obtained. The research uses the cross entropy function of Softmax as the loss function, and the formula is shown in the following (15).

$$L = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})] \quad (15)$$

In (15), y represents the subjective MOS score obtained in the evaluation database, and \hat{y} represents the predicted score. The training iterations are performed according to the loss function to obtain ten classes. After the classification is completed, use 3D-CNN to form a regression prediction model, use transfer learning, load the model parameters saved by classification as the pre-training model of the regression prediction model, and then train the regression model to give the predicted value of the video MOS score. The regression prediction structure of 3D-CNN is similar to the classification structure, the difference is that the fully connected layer of the network structure is followed by a regression prediction node. The schematic diagram of the structure is shown in Fig 4.

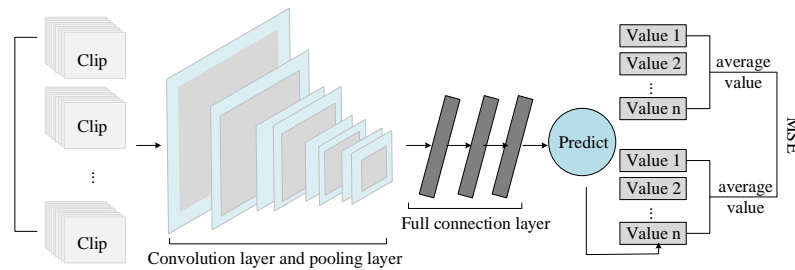


Fig. 4. Prediction structure diagram.

The parameters such as weight parameters and bias items obtained after loading the classification process through transfer learning are loaded to include all convolutional layer and pooling layer parameters, and the parameters of the two fully connected layers are discarded. The loss function uses the mean square error (MSE), and the calculation formula is as follows (16).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2 \quad (16)$$

In (16), N is the number of video clips, y represents the subjective MOS score obtained in the evaluation database, and \hat{y} represents the predicted score. Put the predicted value into the loss function to participate in the regression training, and finally get the most suitable prediction result. Although 3D-CNN can effectively extract features from VR video data, the increase in the convolution dimension will inevitably increase the parameters of the entire neural network, resulting in a sharp increase in the time complexity of the algorithm. The conventional network input is obtained through the calculation output of the upper layer, while the residual network (ResNet) will have a "short-circuit" structure, and the data processed by the multi-layer network and the data processed by the network layer are jointly input and transmitted to in the next layer of the network. The learning goal of ResNet is to solve the residual error, as shown in (17).

$$F(x) = H(x) - x \quad (17)$$

In (17), $H(x)$ is the expected mapping of learning. The difference is obtained through (17), so that the same part before and after the cell mapping highlights the slight changes, so that the deep network structure will not degenerate while ensuring the structural performance. ResNet can reduce the complexity of parameter calculation and reduce the error caused by network depth through the unique design of shortcut. Research on improving 3D-CNN based on ResNet's shortcut. In order to reduce the nonlinear conversion operation and improve the feature extraction ability, a double 3D convolutional layer is designed, and no pooling calculation is performed between the two convolutional layers. Batch regularization calculation is performed after each convolution, and the ReLu function is used as the activation function to reduce the risk of gradient disappearance and gradient explosion. The improved 3D residual convolution unit structure is shown in Fig. 5 below:

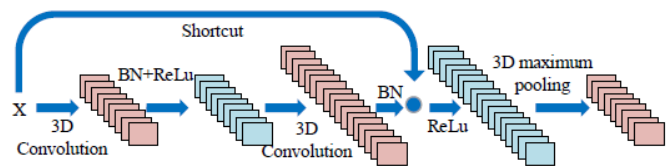


Fig. 5. Structure of 3D residual convolution unit.

Based on the above operations, the study selects video quality evaluation indicators, establishes a subjective evaluation database, uses improved 3D-CNN to classify and predict video data, obtains objective evaluation results, and uses the relationship between subjective evaluation and objective evaluation in the training process. The loss function continuously corrects and trains the model. Finally, a VR video quality evaluation model based on the improved 3D-CNN is obtained, which can accurately evaluate the VR video quality.

IV. ANALYSIS OF VR VIDEO QUALITY EVALUATION MODEL BASED ON 3D-CNN

An accurate and efficient video quality evaluation model is an indispensable tool to measure the pros and cons of video processing algorithms and to control video quality in real time [22]. Therefore, a VR video quality evaluation model based on improved 3D-CNN is constructed. In order to verify the performance of the constructed model, the study used three different data sets to carry out classification training on the research constructed model (model 1), the unimproved 3D-CNN model (model 2), and the BP neural network (model 3), and the convergence of the loss function as shown in Fig. 6 below:

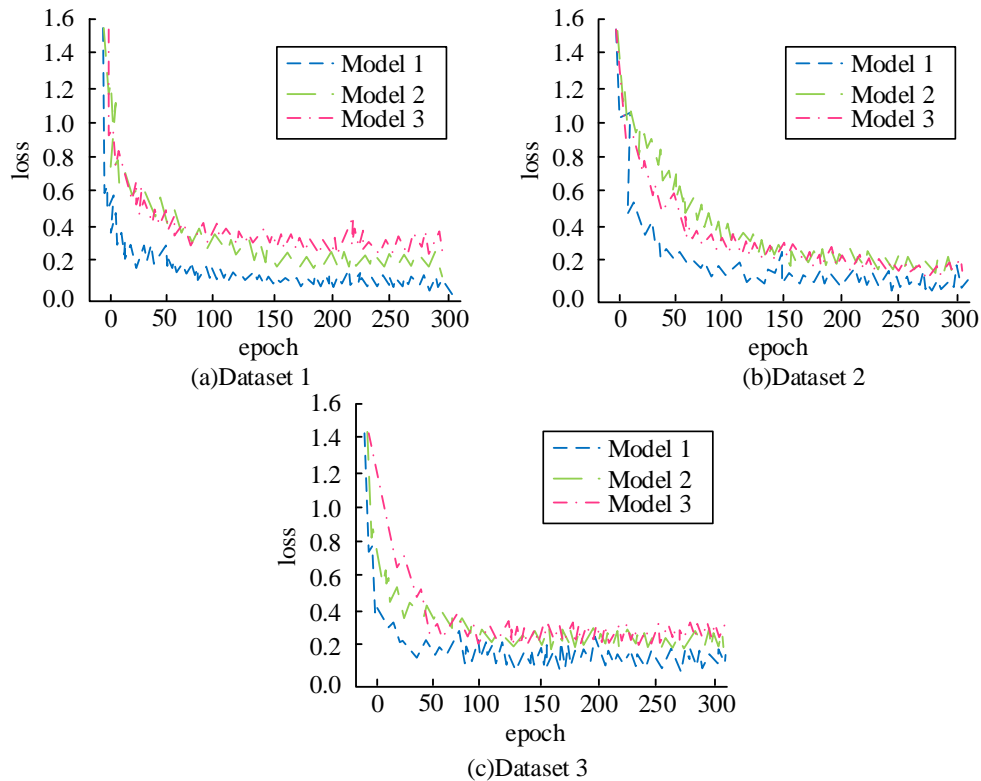


Fig. 6. Comparison of model loss function training.

It can be seen from Fig. 6 that as the number of iterations increase, the value of the loss function of the model decreases, and gradually decreases to a certain extent. Among them, in data set 1, model 1 has the fastest decline rate and reaches the target loss value of 0.16 when the number of iterations reaches 62. However, model 2 reached the target loss value of 0.21 when iterated to 168 times, which was 0.05 higher than model 1. Model 3 reached the target loss value of 0.38 when iterated to 75 times, which was 0.21 higher than model 1; in dataset 2, model 1 reached the target loss value of 0.15 when iterated to 101 times. Model 2 reaches the target loss of 0.28 when iterating to 152 times, which is 0.13 higher than Model 1. Model 3 reaches the target loss of 0.29 after 197 iterations, which is 0.14 higher than that of model 1; in data set 3, model 1 reaches the target loss of 0.11 when iterated to 47 times, and model 2 reaches the target loss of 0.23 when iterated to 128 times. 0.12 higher than Model 1. Model 3 iterates to 149 times

to reach the target loss value of 0.22, which is 0.11 higher than Model 1. Comprehensive comparison and analysis of the content in the above figure, it can be concluded that model 1 has the best convergence effect and the fastest convergence speed.

In order to further verify the improvement effect of 3D-CNN, after the model training, the test set was tested using model 1 and model 2. The results of the linear regression analysis graph obtained from the test and the predicted value and subjective score change fitting graph are shown in Fig. 7 shown.

Comparing Fig. 7(a) (b), we can see that the left picture is a linear regression analysis chart, and we can see the correlation between the predicted score and the subjective score. Before the improvement, the Pearson correlation coefficient of the model was 0.9025, the Spearman coefficient

was 0.8963, and the RMSE value was 0.3258. The Pearson correlation coefficient of the improved model is 0.9437, which is 0.0412 higher than that before the improvement, the Spearman coefficient is 0.9359, which is 0.0396 higher than that before the improvement, and the RMSE value is 0.2051, which is 0.1207 lower; the right picture is the predicted value of the test sample. The degree of deviation from the subjective score. It can be seen that the predicted score curve obtained by the improved model has a high degree of coincidence with the subjective score. Based on the analysis of the content in the

above figure, it can be seen that the improvement of the model can effectively reduce the prediction error.

In order to compare the rationality of introducing viewport cutting into VR video clips of the model more specifically, the MOS score of each evaluation category in the subjective evaluation library before and after cutting is compared with the model prediction results, and the prediction accuracy, missed detection rate and false recognition rate are compared. The details are shown in Table I below.

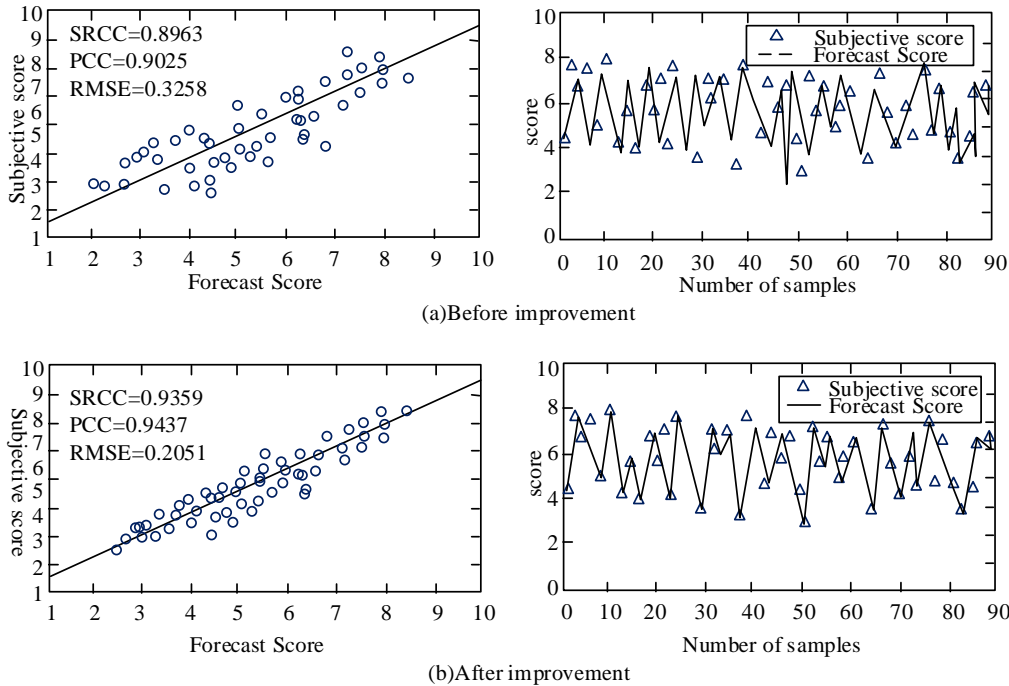


Fig. 7. Model test results before and after improvement.

TABLE I. COMPARISON AND ANALYSIS OF MODEL PREDICTION SCORE AND SUBJECTIVE SCORE

Evaluation type	Before cutting			After cutting		
	Accuracy (%)	Undetected rate (%)	Error rate (%)	Accuracy (%)	Undetected rate (%)	Error rate (%)
Perceived quality	84.10	11.23	15.70	93.26	9.56	7.46
Presence _	83.46	10.96	16.32	94.58	9.32	5.68
Vertigo	86.59	12.36	13.46	93.91	8.99	6.12
Acceptability	87.46	11.02	12.39	95.21	9.12	5.34
Comprehensive _	88.21	11.47	11.04	95.43	8.97	4.69

Comparative analysis of the data in Table I shows that in the process of predicting various subjective evaluation scores, the average prediction accuracy rate of the model trained before the viewport cutting method is 85.97%, the average missed detection rate is 11.41%, and the average misrecognition rate is 85.97%. The average prediction accuracy rate of the model trained after viewport cutting is 94.48%, the average missed detection rate is 9.19%, and the average false recognition rate is 5.86%. Based on the data in the above table, it can be concluded that after viewport cutting,

the number of training samples is increased, and the model can evaluate video quality with higher accuracy.

In order to verify the prediction accuracy of the model and the change of running time under different sample sizes, in addition to the above three training models, the research will also commonly used classification prediction models: logistic regression (Logistics) (model 4), support vector machine (SVM) (model 5). As the number of VR video samples increases, the classification prediction accuracy of the model and the running time are shown in Fig. 8 below:

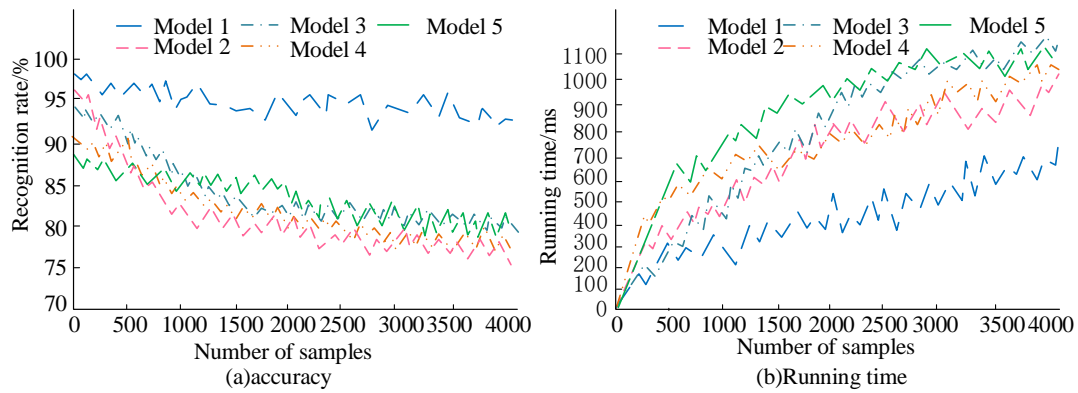


Fig. 8. Variation of prediction accuracy and running time with sample number.

It can be seen from Figure 8 that as the number of sample videos increases, the prediction accuracy of each model decreases and the running time increases. Among them, the accuracy curve of model 1 has the smallest decline. When the sample size is 500, the prediction accuracy of model 1 is 96.43%, and the running time is 0.236s. The prediction accuracy of model 2 is 90.37%, and the running time is 0.347s. The prediction accuracy of model 3 is 93.06%, and the running time is 0.343s. Model 4 has a prediction accuracy of 88.59% and a running time of 0.526s. The prediction accuracy of model 5 is 89.63%, and the running time is 0.623s; when the sample size is 4000, the prediction accuracy of model 1 is 92.64%, and the running time is 0.673s. Model 2 has a prediction accuracy of 76.85% and a running time of 0.921s.

Model 3 has a prediction accuracy of 79.78% and a running time of 1.032s. The prediction accuracy of model 4 is 77.43%, and the running time is 0.996s. The prediction accuracy of model 5 is 80.12%, and the running time is 0.963s. A comprehensive analysis of the content in the above figure shows that model 1 is less affected by the increase in the number of test samples, and the running time and prediction accuracy are in a relatively ideal state.

In order to further verify the classification prediction effect of the model, the study introduces overall accuracy (OA), average accuracy (AA) and Kappa coefficient as evaluation indicators, uses five models to predict three data sets, and compares and analyzes them, as shown in Table II below.

TABLE II. EVALUATION OF MODEL CLASSIFICATION PREDICTION EFFECT

Data No.	Evaluating indicator	model 1	model 2	model 3	model 4	Model 5
Dataset 1	OA (%)	95.86	91.05	89.21	88.01	85.46
	AA (%)	97.90	92.13	88.92	87.64	86.12
	Kappa×100	95.21	91.79	89.64	86.94	85.96
Dataset 2	OA (%)	95.46	90.99	88.54	87.46	86.01
	AA (%)	94.68	91.28	89.46	88.00	85.75
	Kappa×100	96.71	92.03	88.23	87.89	86.23
Dataset 3	OA (%)	94.95	91.78	88.07	87.65	85.69
	AA (%)	95.23	90.89	89.46	87.12	86.49
	Kappa×100	96.62	91.56	88.33	88.04	85.36

Analysis of the data in Table 2 shows that the average OA value of the three data sets in model 1 is 95.27%, the average AA value is 95.94%, and the average Kappa value is 96.18%; the average OA value of the three data sets in model 2 is 90.94%, and the average AA value The value is 91.43%, and the average Kappa value is 91.79%; the average OA value of the three data sets in model 3 is 88.61%, the average AA value is 89.28%, and the average Kappa value is %; the average OA value of the three data sets in model 4 is 87.71%, the average AA value is 87.59%, the average Kappa value is 87.62%; the average OA value of the three data sets of model 5 is 85.72%, the average AA value is 86.12%, and the average Kappa value is 85.85%. Based on the data in the table, the overall accuracy (OA), average accuracy (AA) and Kappa coefficient of model 1 are higher than those of the other four models.

V. RESULTS AND DISCUSSION

With the rapid development of virtual technology and the wide application of various fields, immersive experience with a sense of presence has been widely developed. In order to evaluate the quality of VR video, the improved 3D-CNN was used to construct a VR video quality evaluation model. Through a series of experiments, the following results are obtained. Model 1 has the fastest decline speed and reaches the target loss value of 0.16 when the number of iterations reaches 62, while model 2 and model 3 both need more than 100 iterations to converge. The results show that the proposed model has good convergence performance. After the improvement, the Pearson correlation coefficient increased by 0.0412, Spearman coefficient increased by 0.0396, RMSE value decreased by 0.1207, and the predicted score curve

obtained by the improved model had a high coincidence degree with the subjective score. The results show that the improvement of the model can effectively reduce the prediction error of the model. The average prediction accuracy of the model trained after viewport cutting is 94.48%, the average missing rate is 9.19%, and the average error rate is 5.86%. This shows that the viewport cutting algorithm is reasonable and can effectively improve the model performance. The average OA value of the constructed model was 95.27%, average AA value was 95.94%, and average Kappa value was 96.18%, which could effectively and accurately evaluate the video quality. The number of VR video databases used in this study is limited, and more diverse sample videos can be found for testing and training in subsequent studies to further improve the model.

VI. CONCLUSION

To achieve more precise and accurate VR video quality evaluation, a VR video quality evaluation model based on improved 3D-CNN was constructed. Through the experimental analysis, it is known that the average OA value of the model constructed in the study is 95.27%, the average AA value is 95.94%, and the average Kappa value is 96.18%. It has high heterogeneity with VR video subjective quality score, and the prediction effect is better. Compared with the existing VR video quality evaluation methods, it has a strong competitiveness. In future studies, more diverse sample videos can be found for testing and training to further improve the model.

FUNDINGS

The research is supported by Science and Technology Project of Jiangxi Provincial Department of Education: Research on VR training system of Marathon in 5G era - take the VR training system of Poyang Lake Marathon as an example (No. GJJ202009).

REFERENCES

- [1] M.R. Miller, F. Herrera, H. Jun, et al., "Personal identifiability of user tracking data during observation of 360-degree VR video," *Scientific Reports*, vol. 10, no. 1, pp. 1-10, 2020.
- [2] J. Du, F.R. Yu, G. Lu, et al., "MEC-assisted immersive VR video streaming over terahertz wireless networks: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9517-9529, 2020.
- [3] M.S. Anwar, J. Wang, W. Khan, et al., "Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148084-148099, 2020.
- [4] L. Argyriou, D. Economou, V. Bouki, "Design methodology for 360 immersive video applications: the case study of a cultural heritage virtual tour," *Personal and Ubiquitous Computing*, vol. 24, no. 6, pp. 843-859, 2020.
- [5] M. Teimouri, M. Mokhtarzade, N. Baghdadi, et al., "Fusion of time-series optical and SAR images using 3D convolutional neural networks for crop classification," *Geocarto International*, pp. 1-18, 2022.
- [6] Z. Tu, Y. Wang, N. Birkbeck, et al., "UGC-VQA: Benchmarking blind video quality assessment for user generated content," *IEEE Transactions on Image Processing*, vol. 30, pp. 4449-4464, 2021.
- [7] C. Bonnineau, W. Hamidouche, J. Fournier, et al., "Perceptual quality assessment of HEVC and VVC standards for 8K video," *IEEE Transactions on Broadcasting*, vol. 68, no. 1, pp. 246-253, 2022.
- [8] H. Zhang, L. Dong, G. Gao, et al., "DeepQoE: A multimodal learning framework for video quality of experience (QoE) prediction," *IEEE Transactions on Multimedia*, vol. 22, no. 12, pp. 3210-3223, 2020.
- [9] Y. Tian, H. Zeng, J. Hou, et al., "A light field image quality assessment model based on symmetry and depth features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 5, pp. 2046-2050, 2020.
- [10] G.Y. Lee, S.S. Shin, H. G. Kim, "No-reference sports video-quality assessment using 3D shearlet transform and deep residual neural network," *Journal of Korea Multimedia Society*, vol. 23, no. 12, pp. 1447-1453, 2020.
- [11] J. Yang, T. Liu, B. Jiang, et al., "Panoramic video quality assessment based on non-local spherical CNN," *IEEE Transactions on Multimedia*, vol. 23, pp. 797-809, 2020.
- [12] H. Mzoughi, I. Njeh, A. Wali, et al., "Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 903-915, 2020.
- [13] F. Ramzan, M.U.G. Khan, S. Iqbal, et al., "Volumetric segmentation of brain regions from MRI scans using 3D convolutional neural networks," *IEEE Access*, vol. 8, pp. 103697-103709, 2020.
- [14] J. Liu, K. Zhu, W. Lu, et al., "A lightweight 3D convolutional neural network for deepfake detection," *International Journal of Intelligent Systems*, vol. 36, no. 9, pp. 4990-5004, 2021.
- [15] H. Hassan-Harrirou, C. Zhang, T. Lemmin, "RosENet: improving binding affinity prediction by leveraging molecular mechanisms energies with an ensemble of 3D convolutional neural networks," *Journal of chemical information and modeling*, vol. 60, no. 6, pp. 2791-2802, 2020.
- [16] N. Aldojo, S. Lukas, M. Dewey, et al., "Semi-automatic classification of prostate cancer on multi-parametric MR imaging using a multi-channel 3D convolutional neural network," *European radiationology*, vol. 30, no. 2, pp. 1243-1253, 2020.
- [17] E.S. Salama, R.A. El-Khoribi, M.E. Shoman, et al., "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition," *Egyptian Informatics Journal*, vol. 22, no. 2, pp. 167-176, 2021.
- [18] Y. Tokuoka, T.G. Yamada, D. Mashiko, et al., "3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis," *NPJ systems biology and applications*, vol. 6, no. 1, pp. 1-12, 2020.
- [19] F. Fu, J. Wei, M. Zhang, et al., "Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network," *Nature communications*, vol. 11, no. 1, pp. 1-12, 2020.
- [20] J. Hong, J. Liu, "Rapid estimation of permeability from digital rock using 3D convolutional neural network," *Computational Geosciences*, vol. 24, no. 4, pp. 1523-1539, 2020.
- [21] L. Meng, Y. Tian, S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale," *Journal of applied clinical medical physics*, vol. 21, no. 1, pp. 144-157, 2020.
- [22] D.M. Khan, N. Yahya, N. Kamel, et al., "Automated diagnosis of major depressive disorder using brain effective connectivity and 3D convolutional neural network," *IEEE Access*, vol. 9, no. 1, pp. 8835-8846, 2021.

Design Strategy and Application of Headwear with National Characteristics Based on Information Visualization Technology

Ting Zhang

School of Design and Art, Henan University of Technology
Zhengzhou Henan, 450000, China

Abstract—With the rapid development of big data technology, information technology and visualization technology, traditional national headdress design has gradually been combined with it. The strategies and applications related to national headdress design also fully reflect the beauty of modern science and technology, which is a model of the combination of national classics and modern technology. Based on this, this paper will deeply analyze the various links and processes of the design based on the data based on the specific information of Yao ethnic headwear. At the same time, based on the existing visual design, this paper will take spring, hibernate and other systems as the basic software architecture of the design system and deeply study the visualization principles and data information visualization methods of spring, hibernate and other software, and carry out data information visualization processing on the relevant design of national headwear, to build the corresponding digital material library with national characteristics and the digital design process of national headwear. Through the digital processing and matching of the whole design, the current design of national headwear can be simplified and optimized, and the design efficiency can be improved to provide reference samples for the design of other national characteristics. In the specific design part, this paper will carry out design verification based on Yao nationality's corresponding characteristic headdress design and evaluate the corresponding design from the perspective of artistry, practicality and nationality of headdress design. The practice results show that the information visualization design of national headwear proposed in this paper has obvious advantages over the traditional design, which greatly improves the design efficiency and simplifies the design process.

Keywords—Information visualization; headwear national characteristics; digital material library; yao nationality characteristic headdress design

I. INTRODUCTION

With the rapid iterative development of information technology, visualization technology, as a new technology, is developed and integrated with a large number of disciplines and projects. Traditional information visualization technology processes the data, things, information, models and corresponding knowledge in the design process. It converts them into visual images and videos so that humans can understand through computers [1]. The corresponding data information visualization decision mainly uses visualization technology to transform the important data, decisions, judgments, predictions and corresponding models in the design

process so that they can be seen and understood by humans to provide clear, simple and scientific decision-making reference for decision-makers and designers [2], [3]. Conventional data information visualization technology imports the corresponding data into the preprocessing and processing data module [4], [5]. Then the data module analyzes and models it based on the corresponding data [6], carries out a large number of analogue-to-digital conversions, and carries out visual displays based on this digital information [7], [8] When this kind of software is actually running, it first carries out digital modelling processing for the corresponding processes that need modelling, which mainly includes digital conversion and mathematical statistics of the corresponding data, and then filters the corresponding data based on the statistical results, Through the corresponding query method, we can filter the data that need a visual operation, and finally create the corresponding visual interactive design chart. In the actual design and application, some data may also need the software to carry out the necessary data format conversion, to meet the specific conditions of data visualization [9], [10]. In a word, the development of data information visualization technology is of great significance to realize the scientific, convenient and scientific design process. At the same time, it is also of great significance to promote the corresponding culture and make people better understand the corresponding concepts [11].

Headwear with national characteristics is an important factor in reflecting the national characteristics, culture and customs [12], and it is also an important element to further distinguish this nation from other nations [13], [14]. The headdress with national characteristics reflects the unique aesthetic habits, religious customs and characteristic culture of the nation to a certain extent [15], which is the most distinctive embodiment of the nation. Ethnic minorities pay special attention to the design and dressing of headwear, so headwear also contains a lot of national stories, national changes and development, which itself has very important cultural significance, and it also has important value for the research and promotion of ethnic culture [16], [17]. However, the design of traditional national headwear is often limited to the representative endorsement and teaching of the traditional craftsman of the nation. Its way is too traditional, which makes a large number of craftsmanship techniques and headwear elements gradually lost in historical development, and even some headwear elements lost. Therefore, the design process, design elements and historical changes contained in the

corresponding national headwear Digital preservation of cultural elements are of great significance [18], [19]. Similarly, the personalized design of national headwear based on the national characteristic headwear data information database after complete digital processing is of great significance for the scientific, reasonable and efficient design of national cultural headwear, the inheritance of national cultural development, and the giving of new meaning and new life to national culture [20]. The deep combination of the national headdress design process and information data visualization technology is a model of efficient integration of classic and modern technology.

Visualizations of data that itself can have two-dimensional or three-dimensional semantics have been used for years before computers were used for visualization. Since the computer began to be used in visualization technology, many novel visualization techniques have been discovered and the existing ones have been improved. And the application field extends to large-scale data set visualization and dynamic interactive display. However, for most data stored in databases, there is no standard way to map data to a Cartesian coordinate system because the data has no fixed two-dimensional or three-dimensional semantic properties. In general, a relational database is viewed as a collection of multidimensional attribute data, with each dimensional attribute corresponding to the dimension of the coordinate. If three-dimensional orthogonal coordinates are used to represent visual data, visualization technology, distortion and interaction technology, then all data visualization technology can be considered as a combination of the above three. In this paper, the abstract information in the data table is transformed by visualization technology, that is, the data is stored in the form of visual structure, and the data is represented by multi-dimensional variable values. Different visualization methods will have different visualization structures. Finally, after mapping transformation, the graphical image of the composition is displayed, which also the final result is observed by the user. Each state of the data can be manipulated by the user through human-computer interaction, but the user's action does not change the fundamental structure of each data state.

Based on the above analysis and research, this paper will deeply analyze the various links and processes of the design based on the data and on the specific information of the headwear with Yao characteristics. At the same time, on the basis of the existing visual design, it will take spring, hibernate and other systems as the basic software architecture of the design system and deeply study the visualization principles and data information visualization methods of spring, hibernate and other software, and carry out data information visualization processing on the relevant design of national headwear, to build the corresponding digital material library with national characteristics and the digital design process of national headwear. Through the digital processing and matching of the whole design, the current design of national headwear can be simplified and optimized, and the design efficiency can be improved to provide reference samples for the design of other national characteristics. In the specific design part, this paper will carry out design verification based on Yao nationality's corresponding characteristic headdress design and evaluate the

corresponding design from the perspective of artistry, practicality and nationality of headdress design. The practice results shows that the information visualization design of national headwear proposed in this paper has obvious advantages over the traditional design, which greatly improves the design efficiency and simplifies the design process.

The structure of this article is arranged as follows: the second section of the article will mainly analyze and study the current research status of information visualization technology, national headdress design and other related concepts; The third section of the article will mainly analyze and study the national headdress design based on information visualization technology, focusing on the application of visualization technology and the digital process of national headdress design; The fourth section of the article will take the Yao characteristic headdress design as the object to practice, and compare it with the traditional design; Finally in fifth section the article will be summarized and prospected.

II. RELATED WORKS

With regard to the current situation of data information visualization technology and the design of headwear with national characteristics, the current main research focuses more on the research of isolated data information visualization technology and headwear with national characteristics, while the corresponding fusion research is relatively few. The research on information visualization technology is a hot research topic at present. A large number of scientific research institutions, scholars and universities have conducted a lot of discussion and Analysis on its integration with other disciplines [21]. The current visualization technology is mainly divided into scientific computing visualization technology, data information visualization technology, information visualization technology and scientific knowledge visualization technology. Relevant researchers at Stanford University in the United States have proposed traffic information data visualization technology for the corresponding traffic information technology, established a digital information database based on the corresponding traffic information data, and built a large number of data information models [22], [23]; Relevant travel companies in the United States have established corresponding traffic data information into visual dynamic data charts, realized the analysis and research of travel big data, and carried out research on transportation travel planning based on this analysis [24]; Relevant European researchers have combined visualization technology with medical imaging technology, which realizes three-dimensional imaging of medical images through visualization technology, and has been widely used in diagnosis and treatment planning [25]; Based on this, relevant research institutions in the United States have applied the corresponding visualization technology to the field of communication with medical patients and patients' families. They have built a visualization communication platform based on data visualization technology, which further solves the communication problems of medical problems and further alleviates the problems of doctors and patients [26], [27]. At the corresponding level of national headwear design, there are many relevant studies in this part of the Chinese Mainland. Still, the main research focuses on excavating and protecting the cultural meaning of headwear, and there is little literature to

discuss and analyze its design process. At the same time, the corresponding national headwear design and research are mostly focused on the research and analysis of traditional design methods. At the level of headwear research, Relevant scholars in the Chinese Mainland have analyzed and studied the specific national headdress design, mainly discussing the differences in headdress design behind different ethnic branches and its cultural connotation [28], [29]; Relevant scholars have studied the process and protective significance of headdress design for specific ethnic groups, but have not conducted in-depth research on its combination with modern technology [30]. Therefore, there is relatively little research on integrating information visualization technology and traditional ethnic headdress design, and in-depth research is of great value.

III. RESEARCH METHODOLOGY

This section mainly analyzes and studies the design process of the integration of information visualization technology and national headdress design, including the principle and application of information visualization technology and the national headdress design process based on the Yao nationality. Based on this, the corresponding research principal framework is shown in Fig. 1. It can be seen from the figure that the information visualization technology mainly includes data sorting and analysis, process data statistical analysis, data model establishment, digital data conversion, digital information screening, visual operation and processing, visual presentation and design evaluation. In the corresponding information visualization process, the processes that need to be modelled are first digitally modelled, mainly including the digital transformation and mathematical statistics of the corresponding data. The corresponding data are filtered based on the statistical results, and the data that need to be visually operated are filtered through the corresponding query methods. Finally, the corresponding visual interactive design charts are created. In the actual design and application, some data may also need software to perform necessary data format conversion to meet the specific conditions of data visualization. In the corresponding part of headwear design with national characteristics, it is mainly necessary to carry out data statistics on the relevant element information of headwear, digitally present the corresponding abstract and specific design elements and the corresponding design process, and transform the important data of national headwear, the decision-making of designers, the judgment of designers, the prediction of designers and the corresponding models in the whole design process, so that they can be seen and understood by human beings, To provide clear, simple and scientific decision-making reference for national headdress designers. From the corresponding principal block diagram in Fig. 1, it can be further seen that the design verification analysis in this paper is mainly based on the characteristic headwear of the Yao Minority in the Chinese Mainland. It mainly compares the corresponding design process and design results, design evaluation and traditional design process, and design results and design evaluation proposed in this paper. The main analysis indicators include design efficiency and scientificity [31]. The design method proposed in this paper has obvious advantages from the corresponding design results. In addition,

it can be further seen from Fig. 1 that the visualization software used in this paper is Spring and Hibernate, and its integration application in information visualization technologies and other fields has been widely verified [32].

A. Application Analysis of Information Visualization Technology

Based on the analysis of national headwear and on the information visualization technology proposed in this paper, the corresponding information visualization technology is defined as information visualization interactive design [33]. It is combined with the corresponding characteristics of national headwear design and the corresponding design factors, the corresponding information visualization interactive design technology is divided into five levels, corresponding to the strategic demand level of headwear design, the demand range level of national headwear design, the structure level of national headwear design, the frame layer of national headdress design and the expression layer of national headdress design [34]. Based on this, the information visualization technology flow chart of the corresponding national headdress design is shown in Fig. 2.

In the corresponding strategic demand level of national headwear design, the corresponding needs is to analyze the specific design elements of national headwear in detail, including the cultural heritage, cultural stories, national customs, etc. the corresponding part of the national headwear design is the ultimate goal of this part of the design [35]. For national headwear designers, they can get the corresponding design goals through this part, get the corresponding other design goals based on this design goal, and extend and analyze the economic benefits on the basis of the design goals expected by the designers to make the national headwear design finally have a focus at the beginning of the design.

The range of national headwear design needs level, which is mainly the analysis of national headwear design needs [36]. The corresponding demand objects are different, and the design concepts they face are also different. For the people of this nationality, the corresponding headdress needs to maintain the original flavour, fully reflect the specific cultural characteristics of this nationality, and reflect the cultural connotation and cultural story behind this nationality; For other nationalities, it is necessary to fully meet the curiosity of such people about the national characteristics, meets the combination of national characteristics like aesthetics and modern art, and reflect the artistic beauty of the combination of classics and modernity, to give new historical significance and cultural inheritance to national headwear, and reflect the aesthetic vitality of National headwear in the new era.

The corresponding national headdress design structure layer mainly includes the interaction structure and information structure of the national headdress design. The corresponding national headdress information architecture not only includes the interface display of the front-end design but also includes the information exchange, information storage and display of the background of the design system. In this part, the design system needs to fully respond to the designer's design needs and meet the designer's convenient, fast and scientific concept for national headwear design.

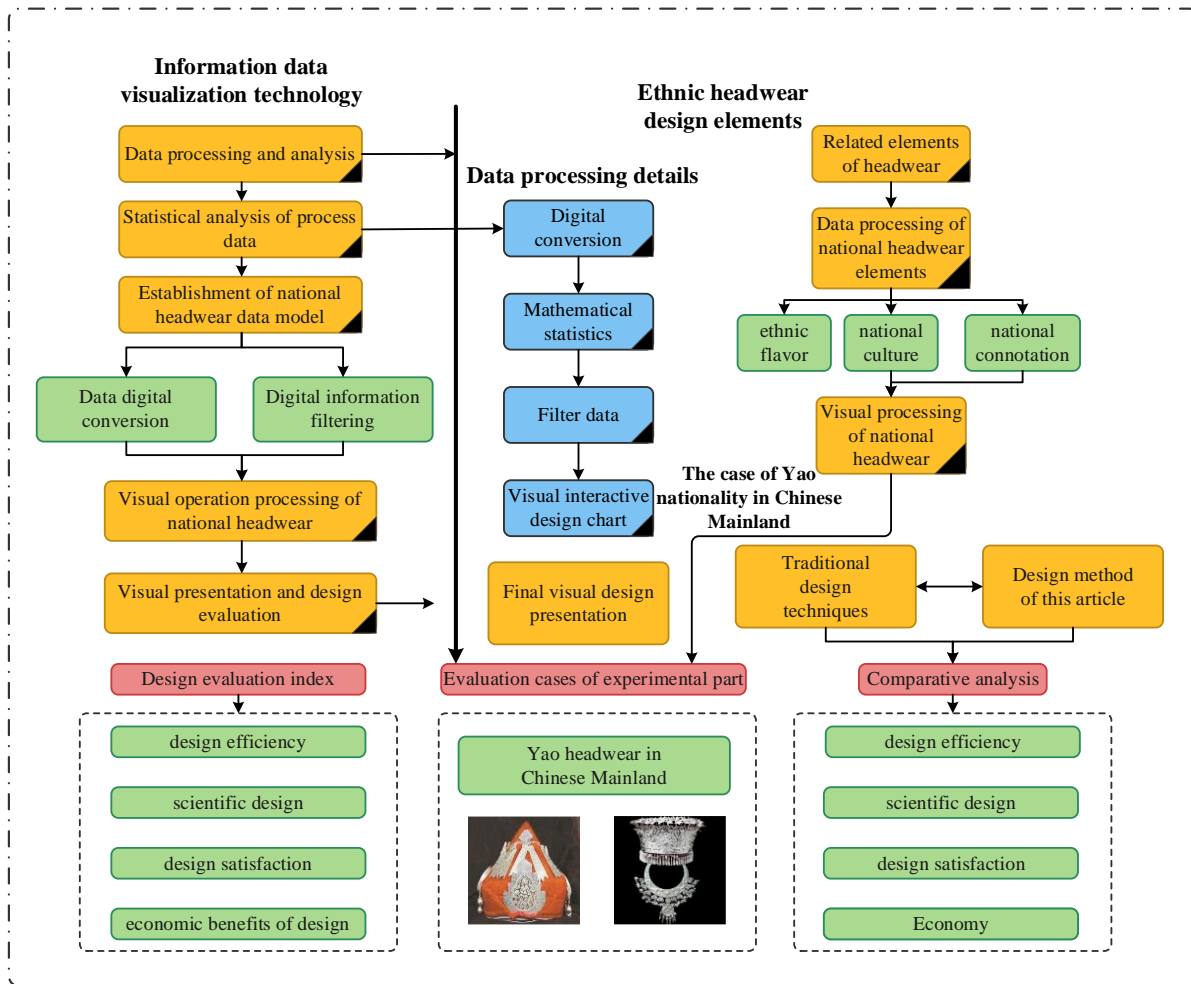


Fig. 1. Principal block diagram of the integration of information visualization technology and national headdress design.

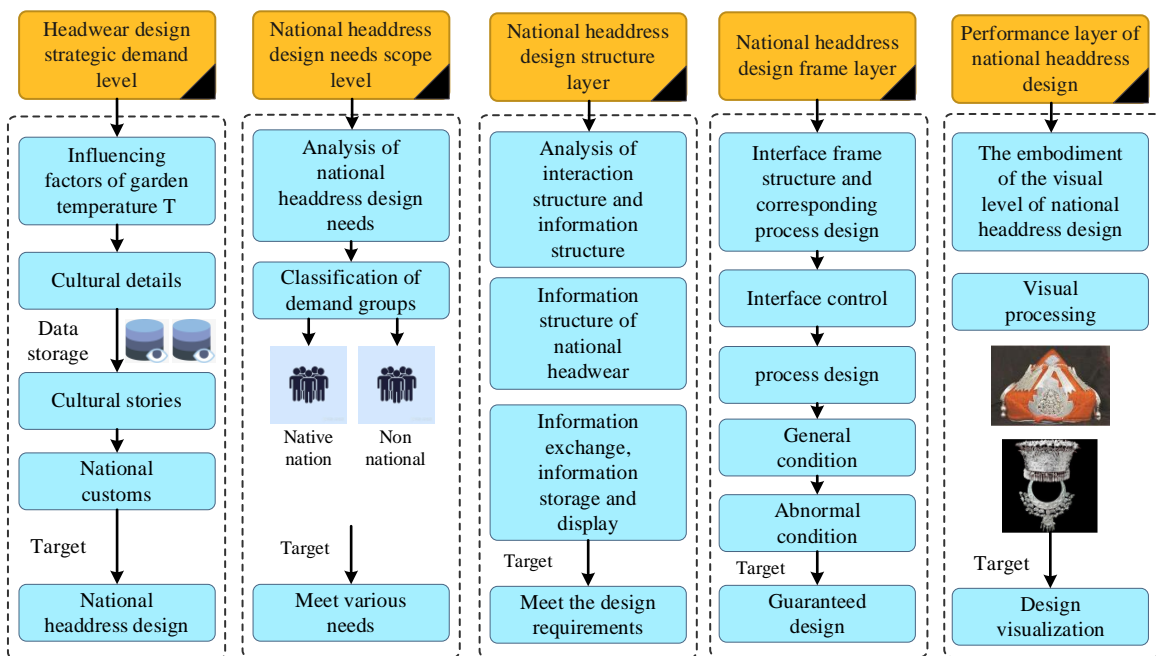


Fig. 2. Design flow chart of information visualization technology for national headdress design.

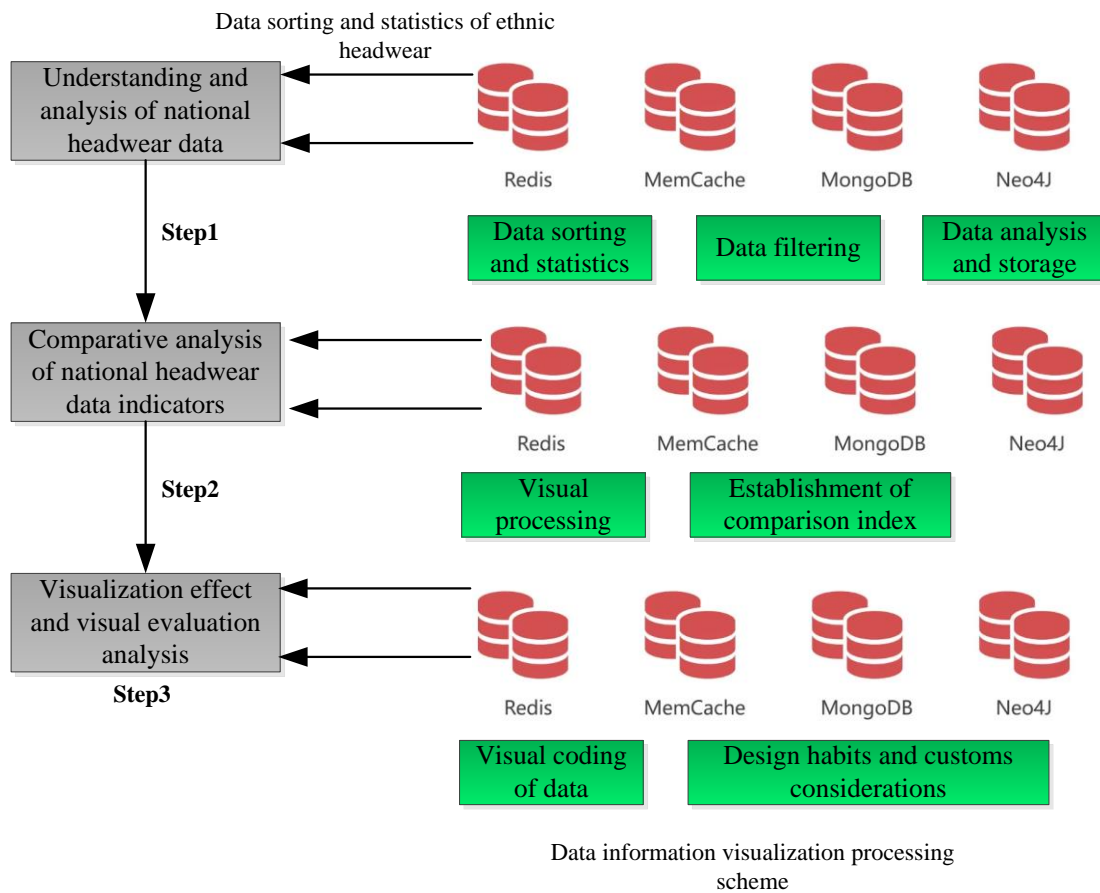


Fig. 3. Design flow chart of data visualization technology for national headdress design.

The corresponding national headwear design framework level mainly includes the interface framework structure of national headwear design and the corresponding process design, including determining various interface control forms, process design, routine condition process and abnormal condition process.

The corresponding performance level of national headdress design, this level mainly focuses on the embodiment of the visual level of national headdress design, that is, the visual presentation of the final headdress design effect [37]. At this level, the designer needs to adjust, analyze and deal with the relevant needs of the demander.

In order to further optimize its visualization processing of national headwear-related data, this paper not only uses information visualization processing but also adopts data visualization technology analysis. The corresponding analysis process is shown in Fig. 3. From the figure, it can be seen that the actual process of national headwear data visualization analysis is mainly divided into national headwear data understanding and analysis, Comparative analysis of national headwear data indicators, as well as visualization effect and visualization evaluation analysis. At the level of understanding and analyzing the corresponding national headwear data, it is mainly to screen, analyze and store a large number of collected data, correctly sort out the relationship between relevant data, and analyze the correctness and scientificity of relevant data;

At the level of comparative analysis of corresponding national headwear data indicators, visualization tools are mainly used to process the corresponding data visually. In this process, appropriate comparative indicators need to be established as a reference; In the corresponding visual data level, it is mainly to encode different data of national headwear visually. In this process, it is necessary to strictly follow the designer's design habits and customs of national headwear.

Based on the above analysis, the relevant data information on national headdress design can realize visualization technology to better provide design services for designers and make the national headdress design process simpler and more scientific.

B. Construction of National Headdress Design System based on Information Visualization Technology

This section mainly realizes the construction of the national headdress design system based on the information data visualization technology, and the corresponding system operation framework is shown in Fig. 4. The corresponding whole system integrates information visualization technology and data visualization technology. It mainly includes a national headdress design performance layer, national headdress design control layer, national headdress design service layer, national headdress design data access object layer and national headdress design database layer. In the corresponding performance layer of national headwear design, it is mainly the

top layer of the system, which is mainly used to collect, sort out and count all kinds of data information and cultural information about national headwear. At the same time, it also needs to interact with other layers to realize the visual processing of national headwear data information, and its corresponding display form is mainly based on visual charts; The corresponding national headdress design control layer is mainly used to process designer design requests and corresponding design process business management [38]. At the same time, data access is realized by accessing the database layer. In this paper, the IOC container in spring is mainly used to manage Dao components and corresponding business logic components; In the corresponding ethnic headwear design service layer, which is mainly located below the control layer, it mainly provides the corresponding design interface for the control layer to call and process, so as to complete the functional design of the ethnic headwear design application module and realize the processing and analysis of the actual design business operation; The corresponding national headwear design data access object layer is mainly used to realize the persistent processing of national headwear data, so that the corresponding design data can be independent of the application program or the business logic of the system design, so that the system data can be easily expanded and run independently; The corresponding national headdress design database layer mainly stores the data in the relational database, which mainly adopts hibernate as the corresponding persistence processing framework.

In addition to the overall design of the system, the database design of the system is equally important. The database data corresponding to the system proposed in this paper comes from the collation and statistics of a large number of design details,

constituent elements, cultural stories and cultural connotations of national headwear. The interface between the corresponding design elements and the design system is processed by web service in this paper. After the actual data sorting and statistics, it is necessary to preprocess and analyze the corresponding headwear information data to obtain the potentially important factors of headwear design in advance. The corresponding data processing process includes the cleaning, screening, selection and corresponding transformation of the original data to ensure the accuracy of the data as much as possible and reduce the time for the designer to query the element data in the actual design. Improve the efficiency of national headdress design. The corresponding data management module mainly includes the system data management module, system data statistics module, system data monitoring module and system data comparison module. The corresponding system data management module mainly realizes the sorting, statistics and distribution of national headwear related data; The corresponding system data statistics module mainly realizes the classification and storage of national headwear related design elements, as well as the storage and memory of corresponding design preferences; The corresponding system data monitoring module is mainly to realize the online monitoring of the use of the corresponding design process data information, so as to avoid some low-level errors and cultural errors at the design level; The corresponding system data comparison module is mainly to compare the final designed finished product with the classic national headdress design, the needs of the demander and other relevant data, so as to assist the designer to complete the design evaluation of the final work, so as to ensure that the designed national headdress not only conforms to the national characteristics of the nation, but also meets the needs of the demander, and improve the satisfaction of the design.

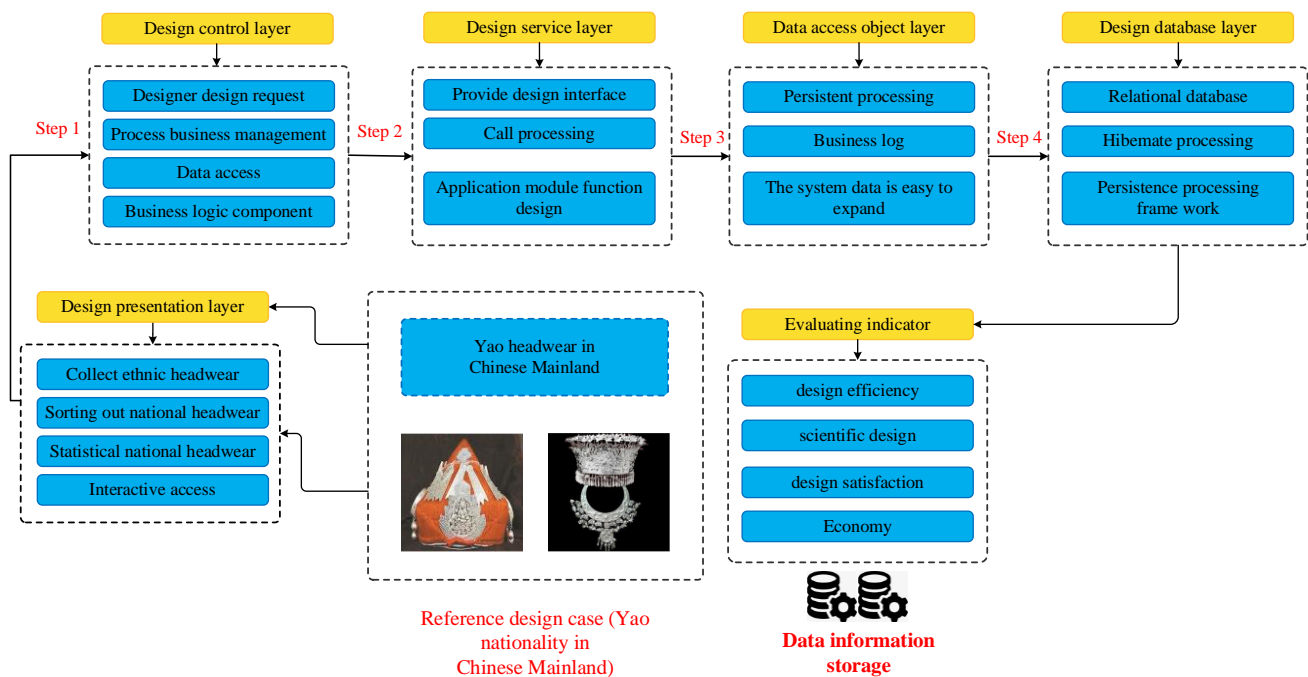


Fig. 4. Frame diagram of national headdress design system based on information data visualization technology.

IV. RESULTS AND DISCUSSION

A. Evaluation Results

To further verify the design advantages of ethnic headwear based on information visualization technology proposed in this paper, this paper selects Yao nationality as a sample for design comparison. The Yao nationality has a long history of national culture, and its headdress design is unique and has rich cultural connotations and aesthetic characteristics. Before the comparative analysis of the actual design, this paper fully collates the pattern, texture, colour, shape and corresponding structural design of the headdress with Yao national characteristics. It makes a statistical analysis of its important elements. Based on the comparative analysis of one of the headwear of the Yao nationality, the e-crown hat, the design process of this paper is compared with the traditional manual design process. The main comparison indicators include the nationality score of the design works, the artistic score of the design works, the practical score of the design works, and the economic score of the design work.

In terms of the nationality of the corresponding design works, questionnaires, visits and other methods are mainly used to show the works under the two design methods, invite the corresponding ethnic and non-ethnic people to score, and take the corresponding average score for evaluation. Based on this, the corresponding evaluation results are shown in Fig. 5. From the figure, it can be seen that the products designed by the design process proposed in this paper have slight advantages in nationality compared with the traditional design

process. The reason is analyzed. It mainly reproduces some complex design patterns lost in the early years.

In terms of the artistry of the corresponding design works, questionnaires, visits and other methods are mainly used to actually display the works under the two design methods, invite the corresponding ethnic, non-ethnic and professional people to score, and take the corresponding average score for evaluation. Based on this, the corresponding evaluation results are shown in Fig. 6. From the figure, it can be seen that the products designed by the design process proposed in this paper have obvious advantages in artistry compared with the traditional design process. Analyze the reasons. The design process proposed in this paper combines aesthetics under modern technology, and its psychological positioning for the demander is more accurate.

In terms of the practicality of the corresponding design works, questionnaires, visits and other methods are mainly used to actually display the works under the two design methods, invite the corresponding ethnic, non-ethnic and professional people to score, and take the corresponding average score for evaluation. Based on this, the corresponding evaluation results are shown in Fig. 7. From the figure, it can be seen that the products designed by the design process proposed in this paper have obvious practical advantages compared with the traditional design process. The reason is analyzed. The design process proposed in this paper combines data visualization technology and information visualization technology, which is more accurate for the demand positioning and processing of the demander.

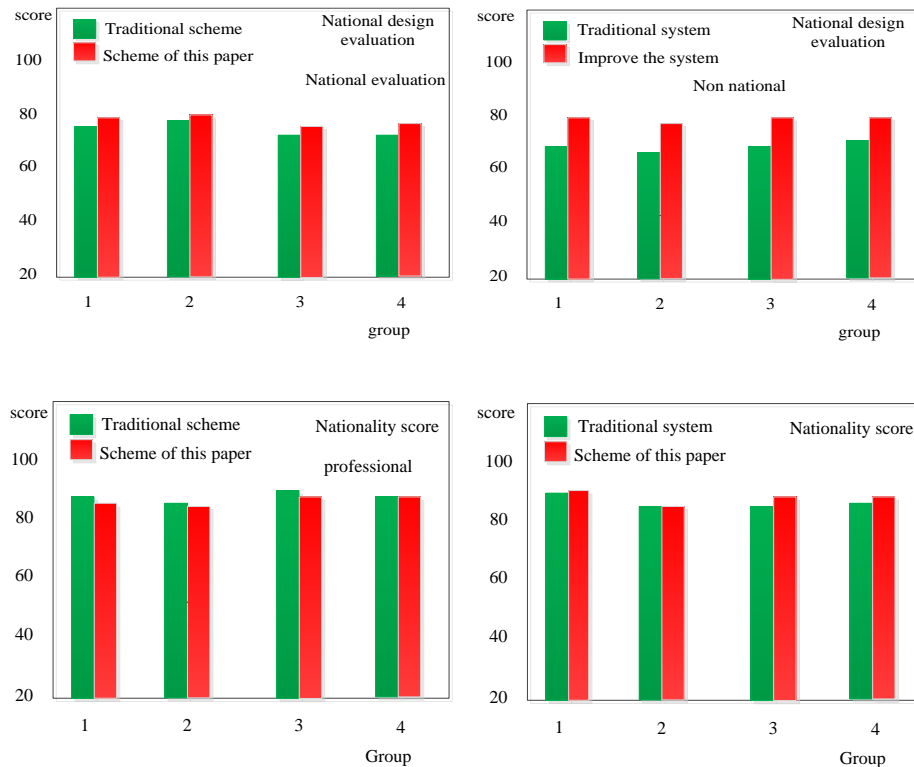


Fig. 5. Comparative analysis of the nationality of design works.

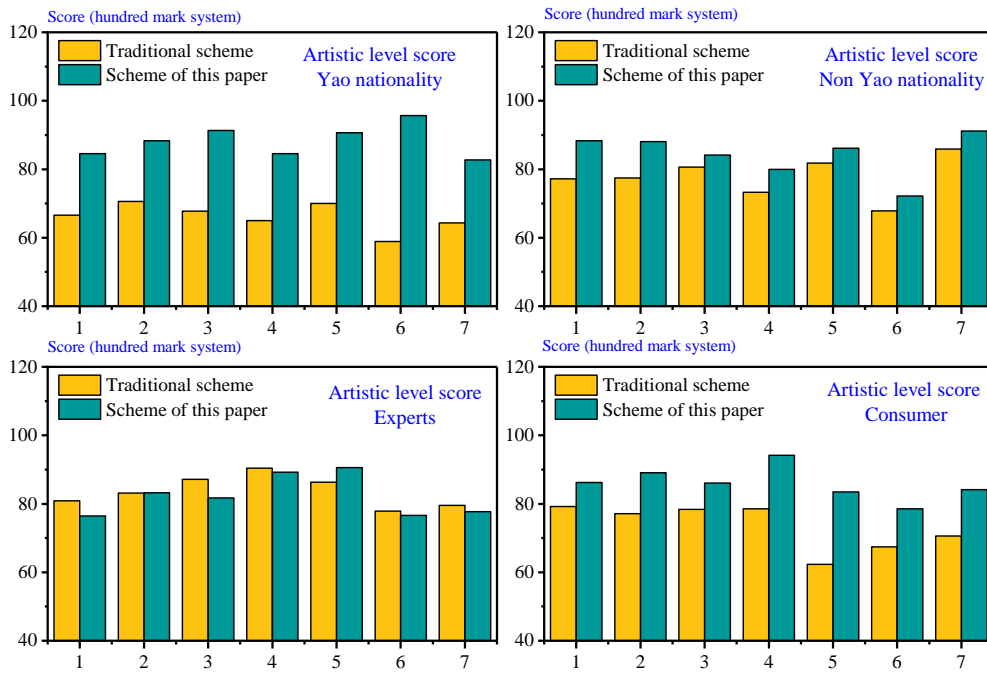


Fig. 6. Comparative analysis of artistic aspects of design works.

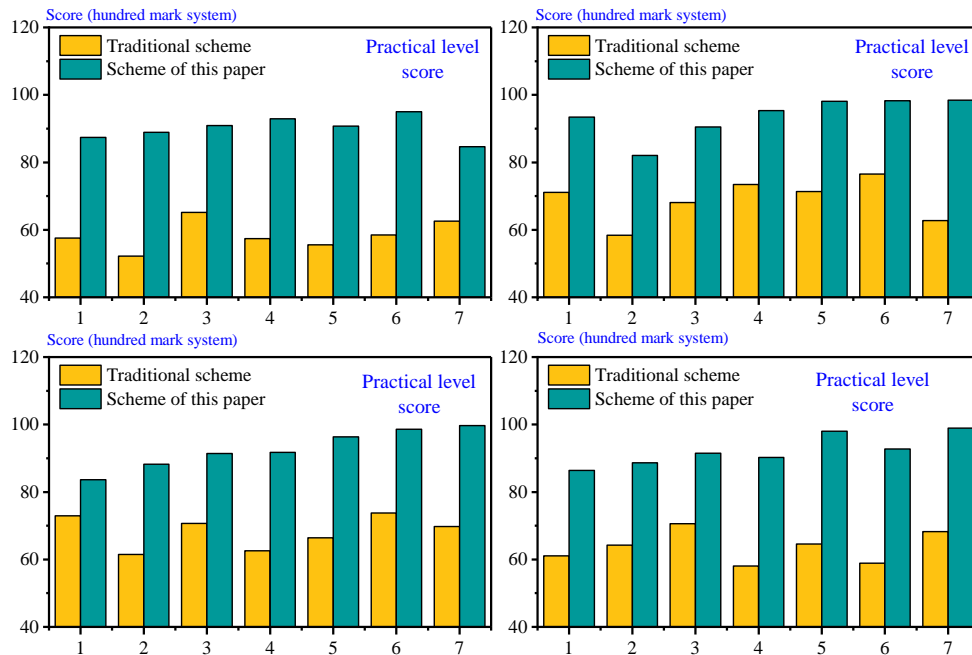


Fig. 7. Comparative analysis of practical aspects of design works.

In terms of the economy of the corresponding design works, questionnaires, visits and other methods are mainly used to actually display the works under the two design methods, invite the corresponding ethnic, non-ethnic and professional people to score, and take the corresponding average score for evaluation. Based on this, the corresponding evaluation results are shown in Fig. 8. From the figure, it can

be seen that the products designed by the design process proposed in this paper have obvious advantages in the economy compared with the traditional design process. The reason is analyzed. The design process proposed in this paper realizes the batch design and production of national headwear design, which greatly saves human design costs and reflects the advantages of modern technology at the design level.

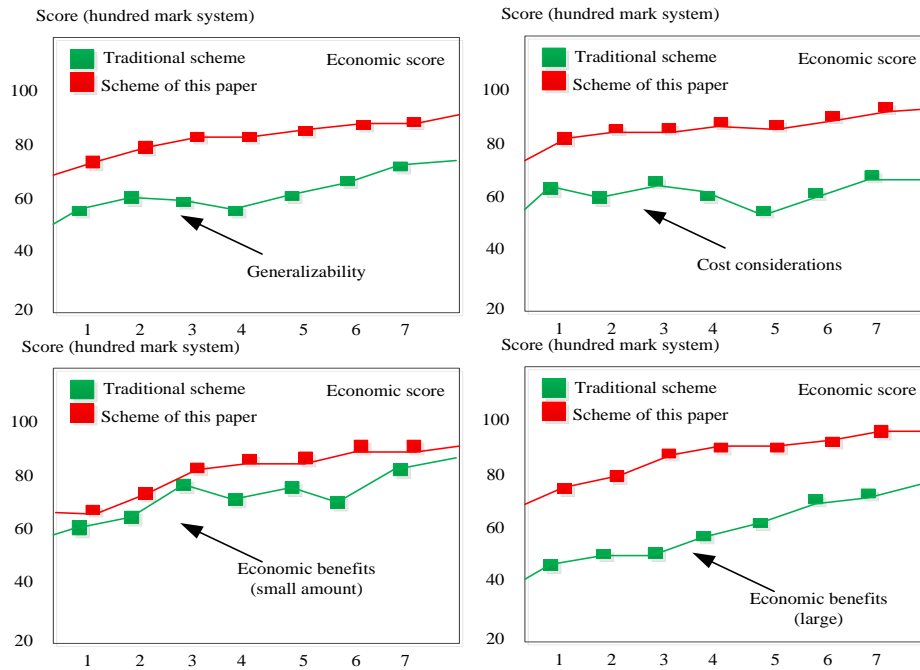


Fig. 8. Comparative analysis of economic aspects of design works.

Based on the above experimental results, it can be found that the national headdress design idea based on information visualization technology proposed in this paper has obvious advantages over traditional design methods, and it is of great significance for further promotion and inheritance of national culture.

B. Discussion

Visualization originates from the promotion and development of computer graphics, user interface and other research fields. Visualization technology refers to the use of computer technology to transform data into graphic form, and to transform abstract things or processes into graphics or images. With the support of computer, information visualization can express abstract data in an interactive visual way, so that people can deepen the cognitive process of data. Visualization technology is widely used in various fields. It presents data in a multidimensional form, which is more conducive to discovering the characteristics of data distribution and its potential semantic relations. Therefore, how to visualize data reasonably in the data space so that users can browse and explore these data conveniently and quickly has become a research hotspot in recent years.

In a computer, information is expressed in a certain structure to facilitate people to store, modify, query, add or delete information and other operations. As a widely used data structure, graph can be used to represent a variety of complex system models. It is basically composed of nodes and edges, where nodes are used to represent the abstraction of entities, and edges connecting nodes are used to represent the relationship between corresponding entities. Uniform, beautiful graphics are very important for understanding and analyzing data. Graph layout algorithm is the basis of graph visualization, and its performance is very important to the effect of data display. At present, most of the algorithms used in graph

visualization draw graphs simply based on the structure of the graph, that is, the connectivity between points. In the personal dataspace system, there are not only structural relations between entities, but also attribute relations determined by their own attributes, so it can also be considered to draw graphs through attribute relations.

National culture is a splendid treasure of China. Each nation has its own unique beauty, which brings us rich and different feelings. The Yao nationality has also shown us its unique national charm, carrying its material and spiritual civilization on its traditional headwear and expressing it with its unique artistic form language. From the shape features of the headwear to the decorative art, all reflect the Yao people's pursuit of beautiful things and longing for the future life. Yao traditional headwear is a very valuable intangible cultural heritage of the motherland. Their wealth is turned into silver ornaments decorated on the headwear. The art of headwear not only has practical application functions, but also is the carrier of Yao civilization. The perfect combination of its function and art is reflected incisively and vividly in the shape and decoration of the headdress.

V. CONCLUSION

Based on the existing visual design, taking spring, hibernate and other systems as the basic software architecture of the design system, this paper deeply studied the visualization principle and data information visualization method of spring, hibernate, and other software and the relevant design of national headwear is visually processed with data information, to build the corresponding digital material library with national characteristics and the digital design process of national headwear. Through the digital processing and matching of the whole design, the current design of national headwear is simplified and optimized, and the design efficiency is

improved to provide reference samples for other national designs. In the specific design part, this paper mainly carries out the design verification based on the corresponding characteristic headdress design of the Yao nationality. It evaluates the corresponding design from the perspective of artistry, practicality and nationality of headdress design. The practice results show that the information visualization design of national headwear proposed in this paper has obvious advantages over the traditional design, which greatly improves the design efficiency and simplifies the design process. In the follow-up research, this paper will pay attention to more national headwear and conduct practical analysis and research based on the design process proposed in this paper. At the same time, this paper will also pay attention to the design process of national clothing and try to combine it with information data visualization technology. In the follow-up research, this paper will also focus on the protection of national culture, design thinking at the level of national culture mining, and realize the processing of corresponding cultural protection and cultural inheritance by collecting more specific data so that the design process proposed in this paper becomes the development of traditional design ideas.

When the amount of data increases, the number of layers increases, resulting in too long edges, which not only makes the drawing space not fully utilized, but also makes the whole graph relatively chaotic. In the future work, we can consider the clustering of nodes and then stratification to reduce the amount of data displayed. In addition, this paper only completed the graph drawing in two-dimensional space. In order to display the graph more aesthetically, the algorithm can be extended to three-dimensional space, and the parameters used in the algorithm can be further optimized to improve the overall efficiency of the algorithm.

REFERENCES

- [1] D. Wilson, E. M. Materón, G. Ibáñez-Redín, R. C. Faria, D. S. Correa, and O. N. Oliveira Jr, "Electrical detection of pathogenic bacteria in food samples using information visualization methods with a sensor based on magnetic nanoparticles functionalized with antimicrobial peptides," *Talanta*, vol. 194, pp. 611–618, 2019.
- [2] G. Shi, X. Shen, Y. He and H. Ren, "Passive Wireless Detection for Ammonia Based on 2.4 GHz Square Carbon Nanotube-Loaded Chipless RFID-Inspired Tag," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-12, Art no. 9510812, 2023.
- [3] Y. He, S. Yang, C.-Y. Chan, L. Chen, and C. Wu, "Visualization analysis of intelligent vehicles research field based on mapping knowledge domain," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 9, pp. 5721–5736, 2020.
- [4] M. Mikusz, S. Clinch, R. Jones, M. Harding, C. Winstanley, and N. Davies, "Repurposing web analytics to support the IoT," *Computer (Long Beach Calif)*, vol. 48, no. 9, pp. 42–49, 2015.
- [5] A. Torres and M. A. Floyd, "Adapted-consumer-technology approach to making near-infrared-reflectography visualization of paintings and murals accessible to a wider audience," *J Chem Educ*, vol. 96, no. 6, pp. 1129–1135, 2019.
- [6] H. Singh and S. Verma, "Visualization of third-level information in latent fingerprints by a new fluorogenic l-tyrosine analogue," *Chemical Communications*, vol. 57, no. 43, pp. 5290–5293, 2021.
- [7] S. L. DeBlasio et al., "Visualization of host-poliiovirus interaction topologies using protein interaction reporter technology," *J Virol*, vol. 90, no. 4, pp. 1973–1987, 2016.
- [8] C. P. Medeiros, M. H. Alencar, and A. T. de Almeida, "Hydrogen pipelines: Enhancing information visualization and statistical tests for global sensitivity analysis when evaluating multidimensional risks to support decision-making," *Int J Hydrogen Energy*, vol. 41, no. 47, pp. 22192–22205, 2016.
- [9] L. M. Raaijmakers et al., "PhosphoPath: visualization of phosphosite-centric dynamics in temporal molecular networks," *J Proteome Res*, vol. 14, no. 10, pp. 4332–4341, 2015.
- [10] G. Shi, X. Shen, F. Xiao and Y. He, "DANTD: A Deep Abnormal Network Traffic Detection Model for Security of Industrial Internet of Things Using High-order Features," *IEEE Internet. Things J.* 2023.
- [11] T. Nauth et al., "Visualization and regulation of translocons in Yersinia type III protein secretion machines during host cell infection," *bioRxiv*, p. 431908, 2018.
- [12] D. M. Berwick, B. James, and M. J. Coye, "Connections between quality measurement and improvement," *Med Care*, pp. I30–I38, 2003.
- [13] J. S. Renzulli, "The national research center on the gifted and talented: The dream, the design, and the destination," *Gifted Child Quarterly*, vol. 35, no. 2, pp. 73–80, 1991.
- [14] H.-W. Suh and H.-S. Ko, "Optimal Stitch Forming Line Identification for Attaching Patches on Non-Developable Clothing Surfaces," *Computer-Aided Design*, vol. 139, p. 103051, 2021.
- [15] P. Cheng, D. Chen, and J. Wang, "Effect of underwear on microclimate heat transfer in clothing based on computational fluid dynamics simulation," *Textile Research Journal*, vol. 90, no. 11–12, pp. 1262–1276, 2020.
- [16] F. De Falco, M. Cocca, M. Avella, and R. C. Thompson, "Microfiber release to water, via laundering, and to air, via everyday use: a comparison between polyester clothing with differing textile parameters," *Environ Sci Technol*, vol. 54, no. 6, pp. 3288–3296, 2020.
- [17] S. He, L. Cheng, W. Xue, Z. Lu, and L. Chen, "An improvement design of groove-wound clothing on the licker-in—Part II. Application on the card machine," *Textile Research Journal*, vol. 89, no. 4, pp. 551–559, 2019.
- [18] Y. Kitase et al., "A new type of swaddling clothing improved development of preterm infants in neonatal intensive care units," *Early Hum Dev*, vol. 112, pp. 25–28, 2017.
- [19] A. Widyanti, M. Mahachandra, H. R. Soetisna, and I. Z. Satalaksana, "Anthropometry of Indonesian Sundanese children and the development of clothing size system for Indonesian Sundanese children aged 6–10 year," *Int J Ind Ergon*, vol. 61, pp. 37–46, 2017.
- [20] Z. Halim and T. Muhammad, "Quantifying and optimizing visualization: An evolutionary computing-based approach," *Inf Sci (N Y)*, vol. 385, pp. 284–313, 2017.
- [21] F. Leite et al., "Visualization, information modeling, and simulation: Grand challenges in the construction industry," *Journal of Computing in Civil Engineering*, vol. 30, no. 6, p. 04016035, 2016.
- [22] A. R. R. de Freitas, P. J. Fleming, and F. G. Guimaraes, "Aggregation trees for visualization and dimension reduction in many-objective optimization," *Inf Sci (N Y)*, vol. 298, pp. 288–314, 2015.
- [23] D. Seo, B. Yoo, and H. Ko, "Responsive geo-referenced content visualization based on a user interest model and level of detail," *International Journal of Geographical Information Science*, vol. 29, no. 8, pp. 1441–1469, 2015.
- [24] J. Seibert, C. W. M. Kay, and J. Huwer, "EXplainistry: Creating documentation, explanations, and animated visualizations of chemistry experiments supported by information and communication technology to help school students understand molecular-level Interactions," *J Chem Educ*, vol. 96, no. 11, pp. 2503–2509, 2019.
- [25] L. Li et al., "Visualization of tumor response to neoadjuvant therapy for rectal carcinoma by nonlinear optical imaging," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 22, no. 3, pp. 158–163, 2015.
- [26] J. Yang and J. Zhu, "Visualization of solids phase separation in a rectangular CFB riser using a novel image calibration method," *Powder Technol*, vol. 273, pp. 76–82, 2015.
- [27] L. Zhao et al., "Using anatomic magnetic resonance image information to enhance visualization and interpretation of functional images: a comparison of methods applied to clinical arterial spin labeling images," *IEEE Trans Med Imaging*, vol. 36, no. 2, pp. 487–496, 2016.

- [28] B. Bialecki, J. Park, and M. Tilkin, "Using object storage technology vs vendor neutral archives for an image data repository infrastructure," *J Digit Imaging*, vol. 29, pp. 460–465, 2016.
- [29] C. D. Prakash and A. Majumdar, "Analyzing the role of national culture on content creation and user engagement on Twitter: The case of Indian Premier League cricket franchises," *Int J Inf Manage*, vol. 57, p. 102268, 2021.
- [30] S. Everley, "The Child Protection in Sport Unit—Supporting national governing bodies in hearing the voices of children: an evaluation of current practice," *Child abuse review*, vol. 29, no. 2, pp. 114–129, 2020.
- [31] P. Reipschlagel, T. Flemisch and R. Dachselt, "Personal augmented reality for information visualization on large interactive displays", *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1182-1192, Feb. 2021.
- [32] M. Cordeil, T. Dwyer, K. Klein, B. Laha, K. Marriott and B. H. Thomas, "Immersive collaborative analysis of network connectivity: CAVE-style or head-mounted display", *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 441-450, Jan. 2017.
- [33] C. Hurter, N. H. Riche, S. M. Drucker, M. Cordeil, R. Alligier and R. Vuillemot, "FiberClay: Sculpting three dimensional trajectories to reveal structural insights", *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 1, pp. 704-714, Jan. 2019.
- [34] L. Besançon, A. Ynnerman, D. F. Keefe, L. Yu and T. Isenberg, "The state of the art of spatial interfaces for 3D visualization", *Comput. Graph. Forum*, vol. 40, no. 1, pp. 293-326, 2021.
- [35] L. Ying et al., "GlyphCreator: Towards example-based automatic generation of circular glyphs", *IEEE Trans. Vis. Comput. Graphics*, vol. 28, no. 1, pp. 400-410, Jan. 2022.
- [36] H. Li, Y. Wang, S. Zhang, Y. Song and H. Qu, "KG4Vis: A knowledge graph-based approach for visualization recommendation", *IEEE Trans. Vis. Comput. Graphics*, vol. 28, no. 1, pp. 195-205, Jan. 2022.
- [37] M. Brehmer, R. Kosara and C. Hull, "Generative design inspiration for glyphs with diatoms", *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 389-399, 2022.
- [38] J. Zeng, Y. Zhao, J. Yu, et al., "Research on the design of virtual image spokesperson of national trendy brand under the concept of meta-universe:--Taking Ling, a national style virtual idol, as an example", *Highlights in Art and Design*, vol. 2, no. 2, pp. 102-107, 2023.

SLAM Mapping Method of Laser Radar for Tobacco Production Line Inspection Robot Based on Improved RBPF

Zhiyuan Liang, Pengtao He, Wenbin Liang, Xiaolei Zhao*, Bin Wei

Liuzhou Cigarette Factory, China Tobacco Guangxi Industrial Co., Ltd., Liuzhou, Guangxi, 545005, China

Abstract—The study focuses on the laser radar SLAM mapping method employed by the tobacco production line inspection robot, utilizing an enhanced RBPF approach. It involves the construction of a well-structured two-dimensional map of the inspection environment for the tobacco production line inspection robot. This construction aims to ensure the seamless execution of inspection tasks along the tobacco production line. The fusion of wheel odometer and IMU data is accomplished using the extended Kalman filter algorithm, wherein the resulting fused odometer motion model and LiDAR observation model jointly serve as the hybrid proposal distribution. In the hybrid proposal distribution, the iterative nearest point method is used to find the sampling particles in the high probability area, and the matching score during particle matching scanning is used as the fitness value, and the Drosophila optimization strategy is used to adjust the particle distribution. Then, the weight of each particle after optimization is solved, and the particles are adaptively resampled according to the size of the weight after solution, and the inspection map of the inspection robot of the tobacco production line is updated according to the updated position and posture information and observation information of the particles of the inspection robot of the tobacco production line. The experimental results show that this method can realize the laser radar SLAM mapping of the tobacco production line inspection robot, and it can build a more ideal two-dimensional map of the inspection environment of the tobacco production line inspection robot with fewer particles. If it is applied to practical work, a more ideal work effect can be achieved.

Keywords—Improved RBPF; tobacco production line; patrol robot; LiDAR; slam mapping; drosophila optimization strategy

I. INTRODUCTION

The inspection robot of tobacco production line can replace manual work to realize remote routine inspection [1]. In case of accidents and special circumstances, it can realize special inspection and customized inspection tasks, and realize remote online monitoring [2]. While reducing manual work, it can greatly improve the content and frequency of operation and maintenance, change the operation and maintenance mode of traditional tobacco production line, realize the intelligent operation and maintenance of tobacco production line, and greatly promote the development of tobacco enterprises [3]. Simultaneous Localization and Mapping (SLAM) technology has always been considered as the most critical step in achieving fully autonomous navigation of robots [4]. As a basic problem of tobacco production line patrol robots, laser

radar SLAM mapping is a prerequisite for tobacco production line patrol robots to achieve other functions. Whether the laser radar SLAM mapping of the tobacco production line patrol robot can be achieved well or not, to a large extent, determines whether the tobacco production line patrol can be successfully completed [5]. Therefore, it is very necessary to study an effective laser radar SLAM mapping method of the tobacco production line patrol robot to help the tobacco production line patrol work. Tobacco production line inspection robots need to obtain real-time environmental information and accurately locate themselves in the production line to effectively perform tasks. The LiDAR SLAM mapping method can realize simultaneous localization and map construction, so it is an ideal choice for tobacco production line inspection robots.

In response to the above problems, many scholars have carried out a lot of effective research, such as the map centered SLAM mapping method of intensive 3D laser anti-radar for tobacco production line patrol robots proposed by Park et al. [6], the active SLAM new method of tobacco production line patrol robots using laser radar to draw maps proposed by Malobick et al. [7], low delay laser radar SLAM for tobacco production line patrol robot using continuous scanning slices proposed by Karimi et al. [8], active SLAM method for tobacco production line patrol robot based on convolutional neural network proposed by Bol et al. [9], SLAM method for tobacco production line patrol robot based on embedded system on chip 3D positioning and mapping researched by Gerlein et al. [10]. Park et al., in their research on the laser radar SLAM mapping method of the tobacco production line inspection robot, took the map as the center, overcome the shortcomings such as the laser radar motion distortion through the local continuous time trajectory, and used the surface resolution maintenance matching algorithm and the surface fusion model based on the normal inverse Wisart to achieve non redundant but dense mapping. The traditional map centered laser radar SLAM method of tobacco production line inspection robot is improved. On the basis of realizing the laser radar SLAM mapping of tobacco production line inspection robot, the accuracy of laser radar SLAM mapping is significantly improved. Malobick et al. used the laser radar as the main sensor in their research on the laser radar SLAM mapping method of the tobacco production line inspection robot. After mapping the inspection environment of the tobacco production line inspection robot, they created a corresponding grid map to navigate the tobacco production line inspection robot system. In Karimi et al.'s research on the laser

radar SLAM mapping method of the tobacco production line inspection robot, 2D Lisaru rotation mode was used to drive the laser radar mounted on the tobacco production line inspection robot, and slice point cloud data from the rotating laser radar was used in the multithreaded matching pipeline for 6D attitude estimation with high update rate and low delay. At the same time, the attitude estimation uses the time motion predictor to better find the feature correspondence in the mapping. The experimental nonlinear optimizer converges quickly, and then completes the attitude fusion operation by using the extended Kalman filter algorithm. Finally, the robot patrol map is updated according to the obtained position and attitude and observation values; In the research of the laser radar SLAM mapping method of the tobacco production line inspection robot by Bol et al., a data set composed of environmental images and the wheel angles related to these environmental images is used to train the convolutional neural network structure so that the convolutional neural network model can learn how to guide the tobacco production line inspection robot. Then, in the inspection environment of tobacco production line inspection robot, it can navigate autonomously through convolutional neural network and obtain corresponding maps at the same time. Gerlein et al. followed a joint design method in their research on the laser radar SLAM mapping method of the tobacco production line inspection robot. By deploying a programmable 3D positioning and mapping chip in the tobacco production line patrol robot system, the laser radar SLAM of the tobacco production line patrol robot is realized, which effectively improves the SLAM mapping efficiency while maintaining a high mapping accuracy. The above methods can realize the laser radar SLAM mapping of tobacco production line inspection robot, but the mapping effect is not ideal.

The improved RBPF (Rao Blackwelled particle filter) is applied to the laser radar SLAM mapping of the tobacco production line inspection robot, which can yield more ideal results of the laser radar SLAM mapping of the tobacco production line inspection robot. Therefore, this paper proposes a laser radar SLAM mapping method for tobacco production line inspection robot based on improved RBPF to better meet the actual work needs. The wheel odometer and IMU data are fused using the extended Kalman filter algorithm, and the fusion odometer motion model and LiDAR observation model are used as the hybrid scheme distribution. In the mixed scheme distribution, the iterative nearest point method is used to find the sampled particles in the high probability region, and the matching score of the particle matching scan is taken as the adaptation value. Then, the Drosophila optimization strategy is used to adjust the particle distribution, and the weight of each optimized particle is solved. According to the weight of the solution, we adaptively do resampling. Finally, the detection map of the production line detection robot is updated according to the updated position and attitude information and the observation information of the tobacco production line detection robot particles.

II. LASER RADAR SLAM MAPPING OF TOBACCO PRODUCTION LINE INSPECTION ROBOT

A. Introduction to RBPF-SLAM Algorithm

In the problem of laser radar SLAM mapping of tobacco production line inspection robot, the theory of probability can be used to mitigate the impact of uncertain factors on the results [11]. Particle filter is not limited by linear Gaussian system, so it can be applied to any nonlinear non Gaussian dynamic system, and has reliable effect in target tracking and positioning [12]. The particle filter RBPF method is applied to the problem of laser radar SLAM mapping of tobacco production line patrol robots. The problem of laser radar SLAM mapping is decomposed into the problem of positioning of tobacco production line patrol robots and the problem of building environmental feature maps based on pose estimation, which can significantly reduce the computational complexity of laser radar SLAM mapping of tobacco production line patrol robots, and has super robustness. It is especially suitable for completing the laser radar SLAM mapping of the tobacco production line inspection robot in small and medium-sized scenes [13].

The core idea of the RBPF-SLAM algorithm is to describe the SLAM problem as a posterior probability of the trajectory in the form of probability [14]. In practical work, the RBPF-SLAM algorithm mainly uses the observation information of the laser radar sensor and the information of the wheel odometer to estimate the environment map m position and posture status of inspection robot in tobacco production line the joint posterior probability of $x_{1:t}$ [15] can be described as $p(x_{1:t}, m | z_{1:t}, u_{1:t-1})$, where, $z_{1:t}$ is marked with observation information obtained by LiDAR, including:

$$z_{1:t} = z_1, z_2, \dots, z_t \quad (1)$$

Among them, z_1 , z_2 as well as z_t refers to each element in the observation information sequence.

$u_{1:t-1}$ is marked with odometer information, including:

$$u_{1:t-1} = u_1, u_2, \dots, u_{t-1} \quad (2)$$

Among them, u_1 , u_2 as well as u_{t-1} are the elements in the odometer information sequence. The RBPF particle filter can be decomposed as follows by using Bayesian formula:

$$p(x_{1:t}, m | z_{1:t}, u_{1:t-1}) = p(m | x_{1:t}, z_{1:t}) p(x_{1:t} | z_{1:t}, u_{1:t-1}) \quad (3)$$

In formula (3), the environment map m trajectory of inspection robot for tobacco production line $x_{1:t}$ the joint posterior probability of is decomposed into the product of two independent posterior probabilities. The motion trajectory of the tobacco production line inspection robot is estimated first, and then the environment map is updated from the motion trajectory combined with the observation data. Among them,

$p(x_{1:t} | z_{1:t}, u_{1:t-1})$ is the posterior probability of the motion path, $p(m | x_{1:t}, z_{1:t})$ is a posteriori probability of the map. Particle filter is required to estimate the potential motion trajectory. At the same time, each particle has a motion trajectory. The final environment map is constructed from the movement tracks of these particles and the observation of the system.

The solution of $p(m | x_{1:t}, z_{1:t})$ is usually based on the extended Kalman filter algorithm, which uses occupancy grid mapping and reverse sensor model to generate planar grid maps, m is divided into a finite number of mesh elements m_i , where each cell contains a value that represents the probability of its being occupied $p(m_i)$, whose values range from "0" to "1". "0" means not occupied, and "1" means fully occupied. The posterior probability density of the map can be approximated as:

$$p(m | x_{1:t}, z_{1:t}) = \prod_i p(m_i | x_{1:t}, z_{1:t}) \quad (4)$$

Among them, i represents the number of particles.

$p(x_{1:t} | z_{1:t}, u_{1:t-1})$ using particle filter algorithm to solve it means that a particle will represent a potential trajectory in a time step and generate a map at the same time. According to Bayesian criteria to have the formula derivation of $p(x_{1:t} | z_{1:t}, u_{1:t-1})$:

$$p(x_{1:t} | z_{1:t}, u_{1:t-1}) = \eta p(z_t | x_t) p(x_t | x_{t-1}, u_t) p(x_{t-1} | z_{1:t-1}, u_{1:t-1}) \quad (5)$$

Among them, η is marked with normalization factor; $p(x_{1:t-1} | z_{1:t-1}, u_{1:t-1})$ represents the trajectory of the tobacco production line inspection robot at the previous time, represented by the particle swarm at the previous time; $p(x_t | x_{t-1}, u_t)$ indicates that the tobacco production line inspection robot is in the $t-1$ Position and posture at every moment x_{t-1} Tobacco production line inspection robot $t-1$ reach t time odometer data u_t when known, the tobacco production line inspection robot t time position and posture the probability distribution of x_t ; $p(z_t | x_t)$ represents the probability distribution of sensor observation data.

In actual work, the environment map of tobacco production line inspection robot is known m with the tobacco production line inspection robot t time position and posture x_t is available

$p(z_t | x_t, m)$ indicates sensor observation data of the probability distribution of z_t . Since many mobile robots are driven by independent translational and rotational speeds, the speed motion model is used for mileage distribution in this paper. After a series of derivation, the trajectory estimation of the tobacco production line inspection robot can be transformed into the corresponding incremental estimation problem. After obtaining the trajectory of the tobacco production line inspection robot, the map is constructed according to the estimated trajectory and observation data.

The RBPF-SLAM algorithm uses the sequential importance resampling filter to estimate the position and pose of the tobacco production line patrol robot and update the environment map. The specific process can be divided into the following four steps:

1) Sampling: as proposed distribution based on motion model $q(x_{1:t}^{(i)} | z_{1:t}, u_{1:t-1})$ sampling, $x_{1:t}^{(i)}$ is the pose of the particle. By particle set $\{x_{t-1}^{(i)}\}$ generate next generation particle sets $\{x_t^{(i)}\}$, where $i = 1, 2, \dots, N$, N is the total number of particles.

2) Weight solving. Resampling according to importance requires solving the weight of each particle $w_t^{(i)}$, the calculation formula is:

$$w_t^{(i)} = \frac{p(x_{1:t} | z_{1:t}, u_{1:t-1})}{q(x_{1:t}^{(i)} | z_{1:t}, u_{1:t-1})} \quad (6)$$

3) Resample. Resampling is performed according to the size of the solved importance weight to form a new particle set. The total number of particles remains unchanged, and each particle has the same weight after resampling N .

4) Map update. Pose of passing particles $x_{1:t}^{(i)}$ and observations $z_{1:t}$ to update the corresponding map $p(m^{(i)} | x_{1:t}^{(i)}, z_{1:t})$.

B. Improve the Laser Radar SLAM Mapping of RBPF Tobacco Production Line Inspection Robot

a) Information Fusion of Odometer and IMU: In this paper, EKF (Extended Kalman Filter) algorithm is used to estimate the position and pose information of mobile robots, and the state model of position and pose fusion is established using wheel odometer and IMU. Since the odometer and IMU data both contain the steering angle and speed information of the mobile robot, the pose state vector of the tobacco production line patrol robot can be described as:

$$x_t = (X_t, Y_t, \theta_t, v_x^t, v_y^t, \omega_t)^T \quad (7)$$

Among them, tobacco production line inspection robot of the position t , attitude and speed information of the time are respectively used (X_t, Y_t, θ_t) , (v'_x, v'_y, ω_t) to mark; T stands for transposition.

If the system status at the moment t is x_t , control input is $(v'_x, v'_y, \omega_t)^T$, and within a certain period of time Δt if it remains unchanged within a certain period of time, then after Δt , the tobacco production line inspection robot $t+1$ the position and posture at all times shall be x_{t+1} . Therefore, the EKF state transfer equation can be obtained as follows:

$$x_{t+1} = x_t + \begin{pmatrix} \Delta t & 0 & 0 \\ 0 & \Delta t & 0 \\ 0 & 0 & \Delta t \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} (v'_x, v'_y, \omega_t)^T \quad (8)$$

Since the odometer data can be obtained directly, the corresponding prediction equation can be obtained as follows:

$$z_{odom,t} = H_{odom} x_t = I_6 (X_t, Y_t, \theta_t, v'_x, v'_y, \omega_t)^T + e_{odom,t}(d) \quad (9)$$

Among them, the predicted value of wheel odometer $z_{odom,t}$ is marked; Prediction matrix of wheel odometer I_t use H_{odom} to mark; $e_{odom,t}(d)$ is the error of the wheel odometer, which obeys the covariance matrix of Gaussian distribution.

In this paper, we only study the environment map in two-dimensional space, so we only need the Z-axis data in the three-axis for IMU, so the prediction equation of IMU can be described as:

$$\begin{aligned} z_{IMU,t} &= H_{IMU,t} x_t \\ &= \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} (X_t, Y_t, \theta_t, v'_x, v'_y, \omega_t)^T + e_{IMU,t} \\ &= \begin{pmatrix} \theta_t \\ \omega_t \end{pmatrix} + e_{IMU,t} \end{aligned} \quad (10)$$

Among them, $z_{IMU,t}$ IMU predicted value is marked; $H_{IMU,t}$ IMU prediction matrix is marked; $e_{IMU,t}$ IMU prediction error is marked, which obeys Gaussian distribution covariance matrix.

The prediction equation of the system can be obtained by combining the prediction equation of the wheel odometer and IMU, which can be described as:

$$z_t = \begin{pmatrix} z_{odom,t} \\ z_{IMU,t} \end{pmatrix} = \begin{pmatrix} H_{odom,t} \\ z_{IMU,t} \end{pmatrix} x_t + \begin{pmatrix} e_{odom,t}(d) \\ e_{IMU,t} \end{pmatrix} \quad (11)$$

By substituting the state transition equation and prediction equation of the position and posture information of the tobacco production line patrol robot into the Kalman filter formula, the position and posture of the tobacco production line patrol robot can be accurately estimated.

b) Distribution of Improvement Proposals: In general, the target distribution is more difficult to obtain an analytical solution than the proposed distribution. Therefore, in practical work, the proposed distribution is often introduced and sampled to approximate the target distribution according to the weight of particles and resampling [16]. The closer the proposed distribution is to the target distribution, the better the effect of particle filter will be [17]. In the RBPF-SLAM algorithm described above, the motion model of the inspection robot in the tobacco production line is regarded as the proposed distribution, and the particle weight is calculated according to the observation model of the inspection robot in the tobacco production line. This method of sampling from the motion model of the inspection robot in the tobacco production line, although relatively simple in calculation, is not accurate. Since the tobacco production line inspection robot will be equipped with a laser radar and a odometer, and the observation accuracy of the laser radar is far higher than the odometer accuracy, as shown in Fig. 1, the probability distribution density function of the odometer model has a wide span and is low and flat, while the probability distribution density function of the laser radar observation model has a small span and is high and sharp. If only the odometer model is used for sampling, a large number of particles will be in meaningless areas. After resampling, these meaningless particles will be discarded. There are fewer particle types in meaningful areas, making the effect of particle filtering worse.

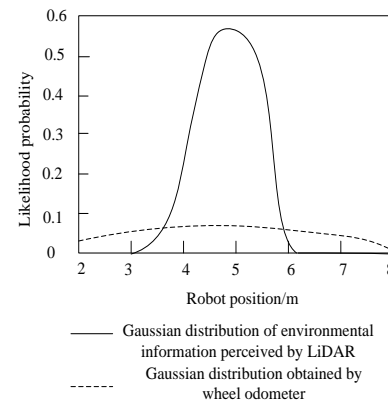


Fig. 1. Legend of distribution density function for LiDAR and odometer measurement models

For this reason, the LiDAR observation value is considered when sampling particles z_t to improve the proposal distribution, and describe the improved proposal distribution as:

$$\Gamma(x_t) = p(z_t | x_t, m^{(i)}) p(x_t | x_{t-1}^{(i)}, u_t) \quad (12)$$

Among them, $\Gamma(x_t)$ is marked the proposed distribution after improvement; $x_{t-1}^{(i)}$ marks the first generation of the previous generation i particles; $m^{(i)}$ is according to the previous generation i current environment map estimated by particles; u_t is according to the previous generation i current odometer information estimated by particles.

After improvement, the proposed distribution comprehensively considers odometer information and LiDAR observation information, and concentrates particles in the high probability interval, which improves the quality of particle set.

c) *Laser Information Scanning Matching*: Considering the characteristics of laser radar scanning, this paper introduces the scanning matching method to calculate the transformation between adjacent laser frames. Scanning matching is to compare two different laser frame information and register them. In other words, it is to find a suitable rotation and translation method, which corresponds the points between different laser frames one by one to get the rotation and translation information of the inspection robot in the tobacco production line. Here, the estimated information of particles is used to register with the laser information, and the analytical formula of the LiDAR distribution is obtained to avoid the error caused by the EKF linearization method.

At present, the commonly used and effective scanning matching method is the iterative closest point method (ICP) [18]. Its principle is the optimal matching based on the least square method, and its main contents include: determining the corresponding relationship of each point cloud data; Calculating transformation matrix; Find the minimum error until the error reaches the set threshold. The specific solution steps are as follows:

1) First, suppose that the laser radar carried by the tobacco production line inspection robot is t the laser data and particle prediction data scanned at any time are L, C , and has:

$$L = \{l_1, l_2, \dots, l_n\} \quad (13)$$

$$C = \{c_1, c_2, \dots, c_n\} \quad (14)$$

Among them, l_1, l_2 as well as l_n are points in the laser data, n marked is the number of points; c_1, c_2 as well as c_n marked are the points in the particle prediction data.

In actual work L as t template point set at time, C as t the set of punctual points to be allocated at the time. About C and L select the point with the smallest Euclidean distance as the corresponding point, and add it to the temporary matching point set.

2) Decentralize each point set to solve L, C and mark their centroids as u_L, u_C , then remove u_L, u_C from L, C .

3) After knowing the corresponding relationship of each point set, solve the objective function shown in equation (15).

$$E(R, T) = \frac{1}{n_C} \sum_{i=1}^{n_C} \|l_i - Rc_i - T\|^2 \quad (15)$$

Among them, $E(R, T)$ the scanning matching error function is marked; For the rotation matrix of point set R to mark; The translation matrix of the point set is used T to mark; l_i, c_i is the corresponding point in the temporary corresponding point set; n_C is marked the number of data points in the particle prediction data set at the current time.

The process of solving the inter frame registration is the process of solving the minimum value of the objective function shown in equation (15), which is completed by using the singular value decomposition method in this paper.

a) *Adaptive Resampling*: Since the RBPF-SLAM algorithm only uses the motion model as the proposed distribution, as time goes on, the cumulative error of the odometer becomes larger and larger, which reduces its estimation performance [19]. For this reason, corresponding improvements have been made in the above sections. When using the motion model to calculate the proposed distribution, the optical observation information has been added, so as to obtain a better proposed distribution. Since the resampling step also has an important impact on particle filtering, during resampling, particles with high weight are usually replaced by particles with low weight. However, frequent resampling operations may eliminate effective particles and cause particle degradation. Therefore, the resampling step is improved adaptively, and a variable to measure particle degradation severity is proposed N_{eff} , which can be described as:

$$N_{eff} = \frac{1}{\sum_{i=1}^N (\tilde{\omega}^{(i)})^2} \quad (16)$$

Wherein, i the weight of particles is marked as $\tilde{\omega}^{(i)}$; For effective particle number is marked as N_{eff} .

In actual work, usually the smaller the value of N_{eff} is, the more serious the particle degradation is. The larger the value is, the better the diversity of particles is. Usually when the N_{eff}

descend to $0.5N$, the resample is performed, and each particle will get the same weight after resampling. Adaptive resampling can resample when the system needs, reduce the number of resampling, and improve the robustness of the algorithm.

b) Drosophila Optimization Strategy: In view of the advantages of the Fruit Fly Optimization Algorithm (FOA) algorithm in target optimization [20], this paper introduces it into the laser radar SLAM mapping work of the tobacco production line inspection robot, and effectively integrates the FOA algorithm and RBPF algorithm to achieve more ideal SLAM mapping work effect.

The basic idea of the fusion is: after the sampling process is optimized and the proposed distribution of sampling particles is obtained, the particle set is regarded as the drosophila population, and the scan matching score is taken as the individual fitness value, which reflects the compliance of the individual with the real state of the tobacco production line inspection robot. The FOA algorithm is used to optimize the particle distribution, drive the drosophila individual to fly to the optimal position, and constantly search the surrounding area for a better position at random, so that the drosophila is constantly approaching the real state of the mobile tobacco production line inspection robot, and alleviate particle degradation.

The standard FOA algorithm regards each drosophila individual as a feasible solution, and constantly moves to the optimal position to seek the global optimal solution of the model. The process of FOA algorithm is as follows:

1) Initialization. Set parameters such as initial position, population size, maximum iteration number Maxgen of *Drosophila melanogaster*.

2) *Drosophila* flies out of the current position and randomly searches for higher concentration positions around. The movement formula is as follows:

$$x_k^{(i)} = x_{k-1}^{(i)} + RandValue \quad (17)$$

Among them, $x_{k-1}^{(i)}$ indicates the particle state at the last iteration; $x_k^{(i)}$ indicates the state of particles at the current time; *RandValue* indicates the step size of random movement.

3) The individual fitness value was obtained from the location concentration of *drosophila melanogaster*, and the optimal individual of the population was found.

4) If the current optimal fitness value is greater than the maximum fitness value of the previous iteration, all *drosophila* flies to the optimal individual. Repeat steps (2) to (4) until the termination conditions are met.

The number of particles determines the number of particles used to represent the state space in the RBPF algorithm. Increasing the number of particles can improve the accuracy and robustness of the algorithm, but at the same time increase the computational complexity. The algorithm usually requires

several iterations to converge to accurate location estimation and map building results. Increasing the number of iterations can improve the stability and accuracy of the results, but it also increases the computation time.

In order to overcome the limitation of the standard FOA algorithm in the convergence process that the population diversity decreases and the particles tend to fall into the local optimum, which leads to the deviation from the real state, this paper introduces the cross mutation operation to improve the adaptability of the population. After the particles are randomly paired, the cross operation is carried out according to the adaptive probability, and then a certain number of optimal particles are copied and mutated to maintain the diversity of the population, prevent falling into local optimal solution. Finally, the state update formula of exponential function step size is used to move *drosophila* individuals to increase the optimization step size, improve the convergence speed of the algorithm, and help the algorithm jump out of the local optimum.

The improved individual renewal formula of *drosophila melanogaster* is:

$$x_k^{(i)} = x_{k-1}^{(i)} + e^{rand} - 1 \quad (18)$$

Among them, e^{rand} represents the step size of exponential function.

By introducing the improved FOA algorithm, the estimation accuracy of the filter can be effectively improved and improved, so that the number of particles required is reduced, thus reducing the calculation amount of the algorithm, and effectively solving the problems of insufficient population diversity and poor real-time performance when completing the laser radar SLAM mapping of the tobacco production line inspection robot based on RBPF.

a) RBPF-SLAM Algorithm Improvement Process

1) The precise motion model and laser radar joint model obtained by fusing the wheel odometer and IMU data are distributed as improvement proposals at the moment $t=0$, select particles N and the weight of particles is N .

2) In the hybrid proposed distribution, the iterative closest point method is used to scan and match to find out the sampling particles in the high probability area, sample in the matched particle set, and take the matching score of the particle matching scan as the fitness value f_i , adjust the particle distribution using the *drosophila* optimization strategy.

3) Calculate, update and normalize the weight of optimized particles, if N_{eff} descend to $0.5N$ start the resample operation.

4) The map is updated according to the position and posture information and observation information of the tobacco production line inspection robot. The improved RBPF-SLAM algorithm flow is shown in Fig. 2.

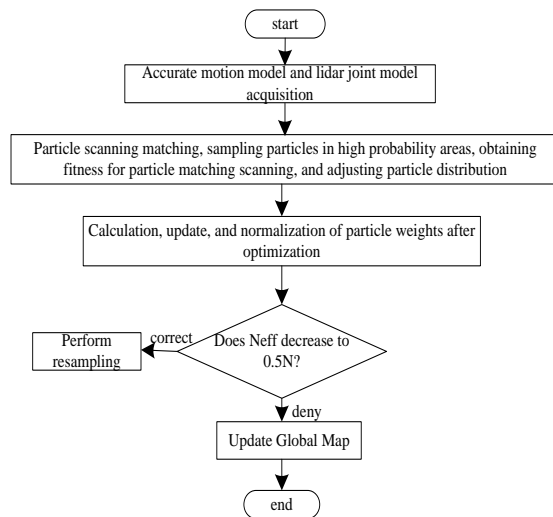


Fig. 2. Improved RBPF-SLAM algorithm process.

III. EXPERIMENT AND ANALYSIS

As an important part of the daily production and management of tobacco production companies, whether the inspection robot of tobacco production line can better realize the laser radar SLAM mapping is directly related to whether the daily inspection task of tobacco production line can be effectively completed. Therefore, in this experiment, the inspection robot of a large tobacco company's production line is taken as the experimental object. The method in this paper is applied to its SLAM mapping, and the effectiveness of the method in this paper is verified. It is reported that the tobacco company was established in June 1984, and two tobacco distribution companies were established in October of the same year. By December 2016, in addition to the original two tobacco distribution companies, it was found that there was a large demand for tobacco marketing in this area, so five tobacco production companies were established in succession. After the establishment of tobacco production companies, it was found that the traditional manual inspection mode was no longer suitable for the long-term development needs of tobacco companies. Therefore, a large number of inspection robots for tobacco production lines were introduced into the newly established five tobacco production branches to assist the production of tobacco production companies and help the development of tobacco enterprises. The introduced tobacco production line inspection robots and their technical parameters are shown in Fig. 3 and Table I. The chassis of the tobacco production line inspection robot uses two wheels for differential movement, and is loaded with wheel odometer, IMU and two-dimensional laser radar. In this experiment, the two groups of tobacco production line patrol robots built in this experiment are respectively 18m * 15m and 24m * 23m in size. In addition, the tobacco production line patrol robots also have a 3D visualization tool RVIZ, which is used to display the laser radar point cloud in the experiment and draw the environment map in real time.

The two-dimensional grid map constructed with 15 sampling particles using the method in this paper is shown in Fig. 4. Among them, the white area is the area that has been scanned by the laser radar, the black line represents the outline

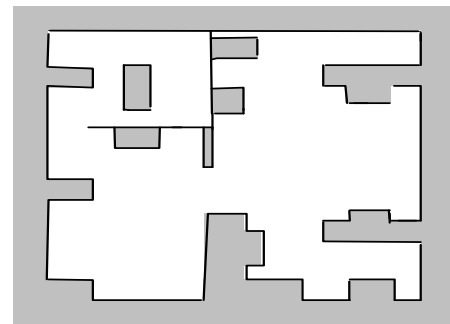
of the object, and the gray area is the map area that has not been scanned by the laser radar.



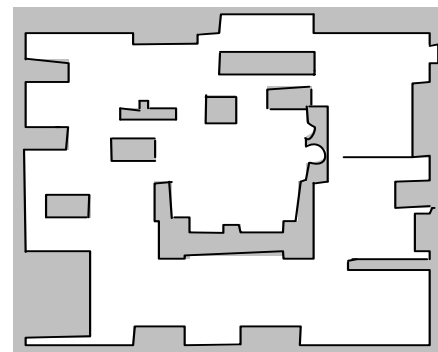
Fig. 3. Tobacco production line inspection robot.

TABLE I. MAIN TECHNICAL PARAMETERS OF TOBACCO PRODUCTION LINE INSPECTION ROBOTS

Parameter information	Parameter value
Model	M-20iA
Overall weight	40kg
Dimensions	800*450*200mm
Gradeability	30°
Fastest running speed	1m/s
Turning radius	2000mm
Visual resolution	1080P
Inspection efficiency	10s/inspection point
Navigation method	Laser SLAM
Repeatability	±10mm



(a) 18m*15m



(b) 24m*23m

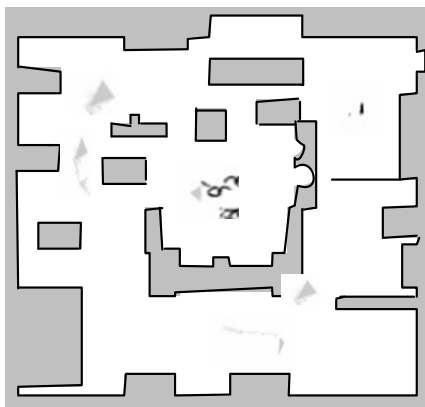
Fig. 4. The method used in this article to construct an environmental map.

It can be seen from Fig. 4 that the method in this paper can realize the laser radar SLAM mapping of the tobacco production line inspection robot, and the environment map constructed is relatively clear, which can better meet the actual work needs. This is mainly because the improved FOA algorithm introduced in the proposed method can effectively improve and increase the estimation accuracy of the filter, thus reducing the number of particles required, thus reducing the calculation amount of the algorithm, and effectively solving the problem of insufficient population diversity and poor real-time performance when the tobacco production line inspection robot based on RBPF completes the LiDAR SLAM mapping.

Fig. 5 is a two-dimensional grid map constructed by the traditional RBPF-SLAM method using 30 sampling particles.



(a) 18m*15m



(b) 24m*23m

Fig. 5. Traditional RBPF-SLAM method for constructing environmental maps.

It can be seen from Fig. 5 that the two-dimensional grid map constructed by the traditional RBPF-SLAM method has a large error, and there are deviations in many parts, such as incomplete scanning in the white area. Compared with the two-dimensional grid map constructed by using 15 sampling particles in Fig. 4, the effect is obviously worse, which is undoubtedly another verification of the effectiveness of this method. This is mainly because the method proposed in this paper uses the exponential function step length state update formula to move fruit flies, increase the optimization step size,

improve the convergence speed of the algorithm, and make the algorithm jump out of the local optimal solution.

In order to further verify the effectiveness of the method in this paper, within 200s, the method in this paper will be reasonably compared with the traditional RBPF-SLAM method in the laser radar SLAM mapping of the tobacco production line inspection robot, and the obtained position and attitude state error comparison curve is shown in Fig. 6.

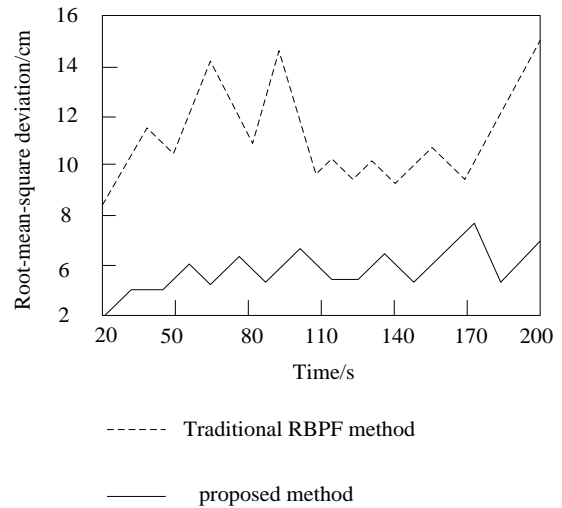
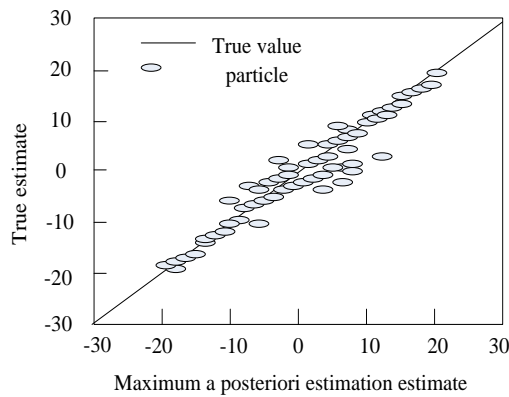


Fig. 6. Comparison curve of pose state error.

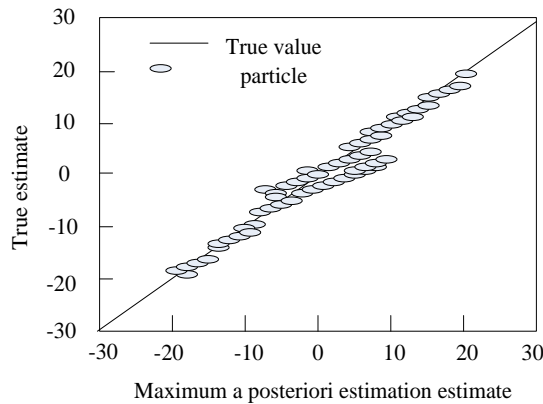
It can be seen from the pose state estimation curve of the tobacco production line patrol robot shown in Fig. 6 that the position state estimation error of the tobacco production line patrol robot obtained by applying the method in this paper to implement laser radar SLAM mapping of the tobacco production line patrol robot is significantly lower than the traditional RBPF-SLAM method. This is mainly because when the method in this paper is used to estimate the position and orientation of the tobacco production line patrol robot, the wheel odometer information and IMU information of the tobacco production line patrol robot are effectively fused, so the position and orientation estimation accuracy of the tobacco production line patrol robot is higher.

Fig. 7 shows the particle distribution after particle filtering of this method and the traditional RBPF-SLAM method.

It can be seen from Fig. 7 that the fitting accuracy of the filtering method in this paper is higher. After filtering, most particles are closely distributed and close to the true value, and there is almost no divergence of particle samples. After traditional RBPF-SLAM filtering, the distribution of particles is relatively divergent, and some particles are far from the true value. Note: The method in this paper has more advantages in the accuracy of map estimation, and can better meet the needs of the actual tobacco production line inspection robot laser radar SLAM mapping work. This improved filtering method has a great advantage in map estimation accuracy, and can well meet the needs of the actual tobacco production line inspection robot LiDAR SLAM mapping.



(a) Particle distribution after filtering using traditional RBPF-SLAM method



(b) Particle distribution after filtering in this method

Fig. 7. Comparison of particle distribution after filtering.

IV. CONCLUSION

The method in this paper can realize the laser radar SLAM mapping of tobacco production line patrol robot, and the mapping effect is good. Its advantages in the laser radar SLAM mapping of tobacco production line patrol robot mainly include:

1) Compared with the traditional RBPF-SLAM method, the method in this paper can build an ideal inspection map of the tobacco production line inspection robot with fewer particles, and the two-dimensional grid map constructed is clearer, which can better meet the actual work needs.

2) Compared with the traditional RBPF-SLAM method, the method in this paper is used to implement the laser radar SLAM mapping for the tobacco production line patrol robot, which can obtain lower position and attitude state estimation error of the tobacco production line patrol robot, and has better fitting accuracy when filtering. After filtering, most particles are closely distributed and close to the true value, and there is almost no divergence of particle samples.

3) The wheel odometer and IMU data are fused using the extended Kalman filter algorithm, and the fusion odometer motion model and LiDAR observation model are adopted as the hybrid scheme distribution. In the mixed scheme distribution, the iterative nearest point method is used to find the sampled particles in the high probability region, and the

matching score of the particle matching scan is taken as the adaptation value.

4) RBPF needs to process a large number of particles to represent the state space, and perform state estimation and map update, which leads to high computational complexity, especially in real-time applications may face the problem of processing time delay. Future research could explore more efficient algorithms, such as improved versions of RBPF based on reducing the number of particles or introducing other optimization methods.

REFERENCES

- [1] Cattaneo, D. , Vaghi, M. , & Valada, A. (2022). Lcdnet: deep loop closure detection and point cloud registration for lidar slam. *IEEE Transactions on Robotics: A publication of the IEEE Robotics and Automation Society*, 38(4), 2074-2093.
- [2] Chunrong, Z. (2021) Bending limit test of robot arm material[J], *Ordance Material Science and Engineering*, 44(4), 120-124.
- [3] Peichao, C. , Kunfeng, L. , & Jiachao, Z. (2023) SLAM and Path Planning for Mobile Robots in Complex Scenes[J], *Computer simulation*, 40(2), 443-448,458.
- [4] Kudriashov, A. , Buratowski, T. , Garus, J. , & Giergiel, M. (2021). 3d environment exploration with slam for autonomous mobile robot control. *WSEAS Transactions on Systems and Control*, 16, 450-456.
- [5] Koval, A. , Karlsson, S. , & Nikolakopoulos, G. (2022). Experimental evaluation of autonomous map-based spot navigation in confined environments. *Biomimetic Intelligence and Robotics*, 2(1), 50-58.
- [6] Park, C. , Moghadam, P. , Williams, J. L. , Kim, S. , Sridharan, S. , & Fookes, C. (2022). Elasticity meets continuous-time: map-centric dense 3d lidar slam. *IEEE Transactions on Robotics: A publication of the IEEE Robotics and Automation Society*, 38(2), 978-997.
- [7] Michal Mihálik, Malobick, B. , Peniak, P. , & Vesteňick, P. (2022). The new method of active slam for mapping using lidar. *Electronics*, 11(7), 1082.
- [8] Karimi, M. , Oelsch, M. , Stengel, O. , Babaian, E. , & Steinbach, E. (2021). Lola-slam: low-latency lidar slam using continuous scan slicing. *IEEE Robotics and Automation Letters*, PP(99), 1-1.
- [9] Bol, N. , Duramaz, M. , Ztrk, E. , Durdu, A. , & Yildiz, B. (2021). Convolutional neural networks based active slam and exploration. *European Journal of Science and Technology*(22), 342-346.
- [10] Gerlein, E. A. , Gabriel Díaz-Guevara, Carrillo, H. , Parra, C. , & Gonzalez, E. (2021). Embedded system-on-chip 3d localization and mapping—esoc-slam. *Electronics*, 10(12), 1378.
- [11] Miller, I. , Cowley, A. , Konkimalla, R. , Skandan, S. , & Kumar, V. (2021). Any way you look at it: semantic crossview localization and mapping with lidar. *IEEE Robotics and Automation Letters*, PP(99), 1-1.
- [12] Sugiura, K. , & Matsutani, H. (2021). A universal lidar slam accelerator system on low-cost fpga, 10(Mar.8), 26931-26947.
- [13] Park, Y. S. , Shin, Y. S. , Kim, J. , & Kim, A. (2021). 3d ego-motion estimation using low-cost mmwave radars via radar velocity factor for pose-graph slam. *IEEE Robotics and Automation Letters*, PP(99), 1-1.
- [14] Lowe, T. , Moghadam, P. , Edwards, E. , & Williams, J. (2021). Canopy density estimation in perennial horticulture crops using 3d spinning lidar slam. *Journal of Field Robotics*, 38(4), 598-618.
- [15] Kim, H. , & Choi, Y. (2021). Location estimation of autonomous driving robot and 3d tunnel mapping in underground mines using pattern matched lidar sequential images. *Journal of Mining Science and Technology: English Edition*, 031(005), 779-788.
- [16] Sung, C. , Jeon, S. , Lim, H. , & Myung, H. (2022). What if there was no revisit? large-scale graph-based slam with traffic sign detection in an hd map using lidar inertial odometry. *Intelligent Service Robotics*, 15(2), 161-170.
- [17] Trybala, P. , John, A. , & Kohler, C. (2022). Towards a mine 3d dense mapping mobile robot: a system design and preliminary accuracy evaluation. *Markscheidewesen*, 129(1) , 18-24.

- [18] Ali, W. , Sheng, L. , & Ahmed, W. (2021). Robot operating system-based slam for a gazebo-simulated turtlebot2 in 2d indoor environment with cartographer algorithm, 15(3), 149-157.
- [19] Belkin, I. , Abramenko, A. , & Yudin, D. (2021). Real-time lidar-based localization of mobile ground robot. *Procedia Computer Science*, 186(14), 440-448.
- [20] Funabiki, N., Morrell, B. , Nash, J. , & Agha-Mohammadi, A. A. (2021). Range-aided pose-graph-based slam: applications of deployable ranging beacons for unknown environment exploration. *IEEE Robotics and Automation Letters*, 6(1), 48-55.

Visual Image Feature Recognition Method for Mobile Robots Based on Machine Vision

Minghe Hu*, Jiancang He

School of Computing, Xinxiang Vocational and Technical College, Xinxiang, China

Abstract—With the continuous advancement of machine vision and computer technology, mobile robots with visual systems have received widespread attention in fields such as industry, agriculture, and services. However, the current methods for processing visual images of mobile robots are difficult to meet the requirements of practical applications. There are issues of low efficiency and low accuracy. Therefore, firstly, spatial information is integrated into the K-means algorithm and image spatial structure constraints are introduced for visual image segmentation. Then the dense connected network is added to the Convolutional neural network structure. This structure is combined with a bidirectional long-term and short-term memory network to achieve visual image feature recognition. The results show that the improved K-means algorithm has a maximum recall rate of 97.35% in the Berkeley image segmentation dataset, with a maximum Randall index of 86.18%. After combining with the proposed improved Convolutional neural network, the highest feature recognition rate for five scenes of mining, risk elimination, agriculture, factory and building is 96.1%, and the lowest error rate is 1.2%. It possesses a high degree of recognition accuracy and is capable of effectively being applied to visual feature recognition on mobile robots, providing a novel reference point for visual image processing on mobile robots.

Keywords—Machine vision; mobile robots; image recognition; convolutional neural network; K-means algorithm

I. INTRODUCTION

Mobile robots play an important role in modern technology, and their environmental perception and decision-making abilities are crucial for achieving autonomous navigation and task execution. Key technologies in a mobile robot's perception system are visual image processing and feature recognition, which provide robots with rich environmental information and accurate target recognition [1-3]. In recent years, the rapid development of deep learning technology has provided new solutions for visual image processing of mobile robots. Traditional Convolutional Neural Networks (CNN), as one of the core algorithms of deep learning, have strong feature extraction and recognition capabilities and have achieved outstanding results in fields such as image classification, object detection, and semantic segmentation. However, applying traditional methods to recognize visual images with mobile robots presents a challenge when using deep neural networks for image processing on the robot due to limited computational resources and power consumption [4-5]. Therefore, the integration of CNN technology with mobile robots for efficient image feature recognition has become a current research focus. Additionally, the K-means clustering algorithm serves as an unsupervised learning method that is widely used for clustering image

features. By grouping image feature vectors, K-means aids in extracting key features of the image, ultimately resulting in image classification and target recognition. However, in the field of mobile robot vision, the application of machine vision is currently limited. Problems persist with low efficiency and accuracy in image feature recognition. In light of this, a study proposes a mobile robot visual image feature recognition method based on CNN and K-means technology. The method incorporates Recurrent Neural Network (RNN) to further improve the image recognition effect of mobile robots.

The paper is divided into four parts. The first part is an overview of the current development status of visual image recognition for mobile robots both domestically and internationally; the second part is to improve the image segmentation technology of K-means clustering and image feature recognition based on CNN and RNN, and constructs a robot image recognition system. The third part is the performance testing and application effectiveness of the system built by the study; the fourth part is a summary statement of the entire study.

II. RELATED WORKS

The CNN and K-means technologies' remarkable ability to identify image features has garnered significant interest from experts. In the context of mobile robot visual image feature identification, the aforementioned technologies are employed to boost accuracy. Jiang et al. proposed a casing infrared fault diagnosis method based on image segmentation and deep learning to effectively distinguish the fault area and background of the casing. During the process, a target detection system was constructed using the CNN framework, and K-means technology was introduced to classify and explore the positions and areas of the obtained images. The data indicates that the algorithm achieves an image classification accuracy rate of up to 98%, exhibiting superior performance [6]. Chen et al. developed a CNN and K-means-based method for scene perception to tackle CNN's low efficiency in different target recognition tasks. This method enabled coarse-grained classification of the input data and lowered complexity in structural design, boosting computational flexibility. Compared with traditional image recognition methods, this method improves accuracy by 36.65% [7]. To solve the time-consuming manual segmentation of brain tumors in magnetic resonance images, Ragupati and Karunakaran combined CNN with K-means to achieve a robust image feature segmentation method. CNN was used to classify images into normal and abnormal ones. Then K-means was used to segment brain tumor images from abnormal brain images. According to the findings, this method

has high accuracy and recognition efficiency [8]. The rapid development of highways and the increase in the number of vehicles require a safe and efficient transportation system for the automotive sector. Therefore, Chen and Zong proposed a license plate recognition model based on CNN-K-means. CNN was used for license plate detection and segmentation, followed by K-means for license plate number detection and segmentation, and finally for recognition. According to the findings, it is more effective and efficient than other models [9]. Rustam et al. proposed an image recognition method based on CNN-K-means to solve the low accuracy in lung cancer image diagnosis. All input data was checked through CNN, and image features were obtained through K-means and transmitted back to CNN for further recognition. The experimental results indicate that the highest performance measurement accuracy is 98.85%. The sensitivity is 98.32%, and the accuracy is 99.40%. This method has good results in lung detection [10].

With the development of society, the visual image feature recognition of mobile robots requires higher precision recognition capabilities. More scholars are researching how to enhance image recognition abilities. Liu et al. believed that estimating the three-dimensional position and direction of objects in the environment using a single RGB camera is very challenging. Therefore, a new neural network module was introduced for detecting three-dimensional objects to achieve the normal movement and operation of mobile robots. The outcomes indicated that mobile robots achieve the most advanced performance [11]. Sungeetha and Sharma found that in the process of 3D image convergence, the projection of different planes was incorrectly recognized in image feature recognition. A machine learning algorithm was proposed as a preprocessing step to enhance the speed and accuracy of data handling. The outcomes indicated that the accuracy is 34.9% higher than that of ordinary digital visual target recognition [12]. Wang et al. used tensor based visual feature recognition methods to identify visual information generated in industrial processes. The research results indicate that this method has good recognition accuracy [13]. Jacob and Darney used DL methods to study user privacy and secure image recognition for IoT management. The research results indicate good performance in improving the appropriateness and robustness of the Internet of Things. Simultaneously, DL greatly improves the accuracy of image feature recognition [14]. Niu et al. used the generating adversarial networks (GANs) to identify defect images in actual production lines. Classical methods face challenges in obtaining adequate defect datasets due to the lack of diversity and quantity. This method repairs defect images through a large number of defect free images on industrial sites. The research results indicate that this method has high accuracy in image recognition. The entire image recognition system has high robustness [15].

In summary, the method for recognizing visual image features in mobile robots using improved CNN and K-means technology has promising applications. The method enhances the environmental perception and target recognition abilities of mobile robots and offers significant support for achieving intelligent navigation and task execution. At present, numerous scholars have researched this problem, but only a few academic achievements have utilized K-means clustering CNN

in recognizing image features, and there exist application limitations. As a result, this paper proposes a method for recognizing mobile robot images that combines CNN and K-means clustering. It aims to offer technical guidance for robot image recognition and to highlight its potential in future research and application.

III. ROBOT VISION IMAGE FEATURE RECOGNITION BASED ON MACHINE VISION

This chapter consists of two parts. The first section improves the K-means algorithm to complete visual image segmentation. In the second section, CNN and RNN are combined to carry out image feature recognition.

A. Image Segmentation Based on K-Means Clustering

K-means is one of unsupervised learning, which does not need to provide label information. It has a very concise objective function. The operation process of the K-means algorithm mainly contains four steps. The first is to randomly initialize the cluster center. This step randomly obtains K points corresponding to the center of each class of clusters. The second step is to calculate the distance from the sample point to the corresponding center. By comparison, the center with the smallest distance is selected. This means that the sample points and their corresponding centers are all of the same class of clusters [16-17]. Based on the clusters created in the second step, the third step recalculates the centers of all the clusters. The process of the second and third steps is then repeated until the termination conditions are met, concluding the algorithm. Fig. 1 illustrates the operation process of the K-means.

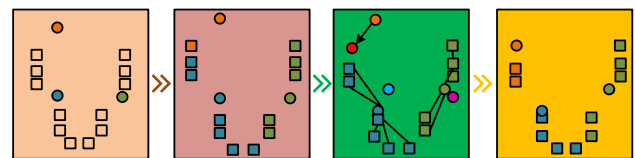


Fig. 1. Illustrate the operation process of the K-means.

If a dataset is $X = \{x_1, x_2, \dots, x_n\}$, the sample in that dataset is x_n . Usually, Euclidean distance is used to partition class clusters. The closest ones belong to the same cluster, while the points of different clusters are farther apart. The objective function of this algorithm is obtained as shown in Equation (1).

$$\min_{\gamma_{nk}, \mu_k} \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} \|X_n - \mu_k\|^2 \quad (1)$$

In Equation (1), z is the samples in the dataset. E is the clusters. μ_k is the center of the k -th cluster. $\gamma_{nk} \in \{0,1\}$ represents whether the n -th sample point is in the k -th cluster. Among them, 0 indicates that the sample is not in the cluster, and 1 indicates that it is in the cluster. Meanwhile, each sample is only in a unique cluster. The K-means algorithm's objective function is to obtain the sum of squared errors. The goal is to minimize this error for each class cluster. This can achieve maximum compactness of cluster samples and ensure maximum distance between cluster samples. The Expectation-maximization is used to address the K-means algorithm. This algorithm is an efficient heuristic algorithm. It can ensure that

the algorithm converges to a local optimal solution in an extremely short time [18]. The Expectation–maximization algorithm is used to solve the objective function of K-means clustering. The first step is to take the derivative. The obtained content is shown in Equation (2).

$$-2 \sum_{n=1}^N \gamma_{nk} (x_n - \mu_k) = 0 \quad (2)$$

The simplified expression of μ_k is shown in Equation (3).

$$\mu_k = \frac{\sum_{n=1}^N \gamma_{nk} x_n}{\sum_{n=1}^N \gamma_{nk}} \quad (3)$$

The specific values of all centers can be obtained through Equation (3). The cluster to which the sample point belongs can be reassigned, as shown in Equation (4).

$$\gamma_{nk} = \begin{cases} 1, k = \arg \min_k \|x_n - \mu_k\|^2 \\ 0 \end{cases} \quad (4)$$

According to the solving process of the K-means in Equations (2) to (4), although it is difficult to ensure a global optimal solution, this algorithm can usually achieve the expected experimental objectives. When segmenting images with K-means, the clustering condition requires the presence of pixels with similar values. If such pixels are absent, the image belongs to a different cluster. Therefore, structural constraint information is relatively lacking. To address this issue, a K-means method for image spatial structure constraints was designed. This method adds image spatial structure constraints on the basis of the original algorithm. The color features and pixel position constraints in the image segmentation process are considered together to improve the segmentation effect [19]. The proposed K-means clustering model for image spatial structure constraints is shown in Fig. 2.

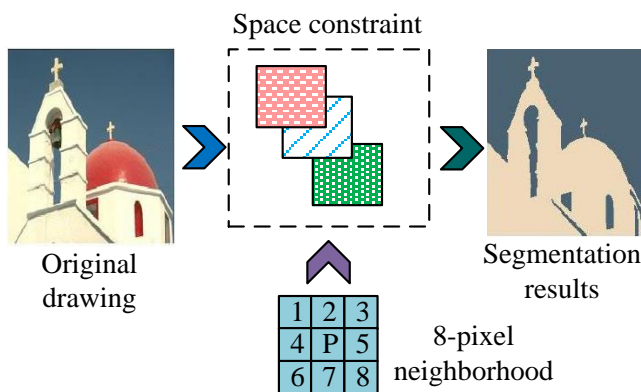


Fig. 2. Schematic diagram of the proposed K-means clustering model for image spatial structure constraints.

In Fig. 2, image segmentation is performed using the K-means algorithm, with sample points being the corresponding pixels in the color image. The clustering is completed under the specified number of clusters. Subsequently, the clustering

center is used to replace the corresponding pixel points to achieve image reconstruction. Under the constraint of image spatial structure, K-means combines itself with spatial structure constraint information. The penalty constraint term corresponding to adjacent pixel points is added to the objective function. The new objective function is calculated, as shown in Equation (5).

$$\min_{\gamma_{nk}, \mu_k} \sum_{n=1}^N \sum_{k=1}^K (\gamma_{nk} \|x_n - \mu_k\|^2 + \alpha \sum_{p=1}^P |\gamma_{nk} - \gamma_{pk}|) \quad (5)$$

In Equation (5), α greater than 0 is a hyperparameter. The main function is to balance the important relationship between the structural constraint term and the reconstruction error term. $KL(\bullet)$ is the neighborhood of the sample points, as shown in Fig. 3. It can be observed that for the internal, boundary, and corner pixels, the corresponding neighborhood pixels of the sample are 3, 5, and 8, respectively. The K-means algorithm for image spatial structure constraints adds the proposed spatial constraint term. As a result, similar color features can affect clustering performance. The position constraint relationship corresponding to adjacent pixel points is taken into account, greatly increasing the persuasiveness of image segmentation results.

A	1							
3	2						1	2
							3	B
							4	5
			1	2	3			
			4	C	5			
			6	7	8			

Fig. 3. Number of neighborhoods of sample points.

Then, the Expectation–maximization is used to address the proposed K-means objective function. The simplified expression of γ_{nk} is shown in Equation (6).

$$\gamma_{nk} = \begin{cases} 1, k = \arg \min_k \|x_n - \mu_k\|^2 + \alpha \sum_{p=1}^P |\gamma_{nk} - \gamma_{pk}| \\ 0 \end{cases} \quad (6)$$

B. Image Feature Recognition Based on CNN and RNN

After completing image segmentation using the proposed K-means, CNN is further applied for image feature recognition. CNN is an adaptive abstract feature extraction model. The structure consists of an input layer, pooling layer, convolutional layer (CL), fully connected layer, and output layer. Among them, the pooling layer and CL connect adjacent nodes through sparse connections, which have the advantage of adaptive feature data extraction [13]. Fig. 4 illustrates the specific structure of CNN.

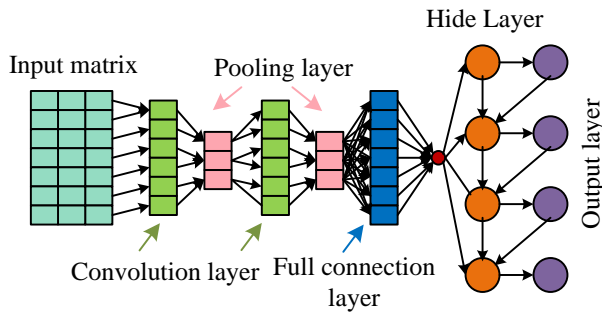


Fig. 4. CNN structure.

In the structure of CNN, the input layer receives preprocessed network data through appropriate dimensions. The convolution layer mainly performs convolution operations on the input values of the Receptive field. The pooling layer is responsible for dimensionality reduction of feature data and filtering out excess information. The fully connected layer combines the local abstract features extracted from the convolutional and pooling layers and reflects them into the label space. The output layer is responsible for outputting the prediction results of the network. Among them, the convolution operation is shown in Equation (7).

$$X_j^{(l)} = f \left(B_j^{(l)} + \sum_{i \in N_l} W_{ij}^{(l)} \times X_i^{(l-1)} \right) \quad (7)$$

In Equation (7), $X_j^{(l)}$ represents the j -th feature output of the l -th CL. N_l represents the set of inputs from layer l . $X_i^{(l-1)}$ refers to the data extracted by the convolutional kernel. $W_{ij}^{(l)}$ stands for the weight of the convolutional kernel. $B_j^{(l)}$ is the bias term. The increase of network layer in DL models is beneficial for feature extraction. However, excessive network layers can lead to overly complex model parameters, making it difficult for errors to be transmitted through gradient backpropagation. Therefore, based on the CNN structure, the DenseNet structural model is introduced. This model can extract abstract features at different levels and merge them. Feature information can be utilized to the greatest extent possible. At the same time, each CL has a fast channel connecting the input and output layers, making it easier for errors to update network parameters through gradient backpropagation. The DenseNet structural model is shown in Fig. 5.

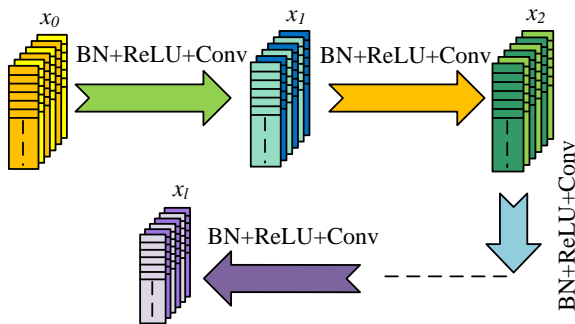


Fig. 5. DenseNet model structure.

In the DenseNet structure, when the CL is L , the connection channels between layers are $L(L+1)/2$. These channels are mainly used for the transmission of feature information. All previous layer output features are combined to complete each subsequent layer input. The calculation is shown in Equation (8).

$$x_l = H_l \left([x_0, x_1, \dots, x_{l-1}] \right) \quad (8)$$

In Equation (8), x_0 represents the input data of the first layer. $[x_0, x_1, \dots, x_{l-1}]$ is the combination of output data from layer l to layer $l-1$. H_l refers to the feature extraction and transformation of the layer, including ReLU Activation function, standardization and convolution operations.

RNN is a special model with "memory" function. The nodes in its hidden layer can receive both the current input signal and the previous output signal. Therefore, the current state information is determined by the hidden layer node input and the previous node output. The calculation is shown in Equation (9).

$$p^{(n)} = q + Ux^{(n)} + Wh^{(n-1)} \quad (9)$$

In Equation (9), $p^{(n)}$ represents the intermediate variable. $h^{(n)}$ refers to the hidden state of node n . W and U represents the weights of the previous hidden state and the current input, respectively. $x^{(n)}$ is the current input information. q is the offset term for the hidden layer. The calculation of $h^{(n)}$ is shown in Equation (10).

$$h^{(n)} = f(p^{(n)}) \quad (10)$$

In Equation (10), f stands for the Activation function. The state information value of the hidden layer is shown in Equation (11).

$$o^{(n)} = c + Vh^{(n)} \quad (11)$$

In Equation (11), $o^{(n)}$ is the node output. V represents the weight of the output. c represents the bias term. The target value corresponding to $o^{(n)}$ is shown in Equation (12).

$$y^{(n)} = \text{Soft max}(o^{(n)}) \quad (12)$$

In Equation (12), $y^{(n)}$ refers to the target value of $o^{(n)}$ mapped to the probability space through the Softmax function. RNN networks have significant advantages in processing sequence information. The traditional LSTM, in contrast, only facilitates one-way memory, relying on previous information to predict output results. It cannot meet the encoding requirements in reverse order. Therefore, a Bidirectional Long Short-term Memory (BLSTM) is introduced, which can model from front to back and from back to front, and output results based on contextual information. If the length of the input sequence is T , the structural model of BLSTM is shown in Fig. 6.

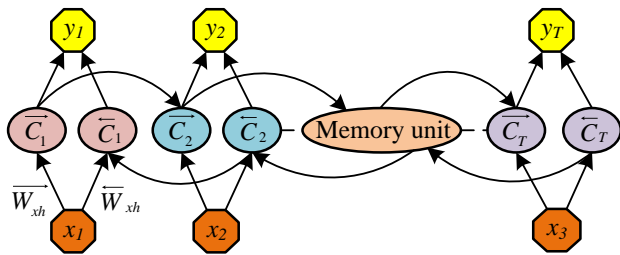


Fig. 6. BLSTM network structure.

In the BLSTM network, the memory units in the forward layer iterate from time step 1 to T, as shown in Equation (13).

$$C_n = H(q_c + W_c g C_{n-1} + W_x g x_n) \quad (13)$$

In Equation (13), C_{n-1} and C_n respectively represent the previous and current state of the unit. x_n represents the current input data. The backward layer transmits information from the end to the front. The calculation is shown in Equation (14).

$$C_n = H(q_c + W_c g C_{n+1} + W_x g x_n) \quad (14)$$

In Equation (14), C_{n+1} refers to the next state of the unit. The output is determined by the forward and backward cell propagation. The output is shown in Equation (15).

$$y_n = q_y + W_y g C_n + W_y g C_n \quad (15)$$

In Equation (15), y_n represents the output of time step n . W_y and W_y refer to the output weight. q_y represents the bias term. Compared to unidirectional RNN, BLSTM structure can extract more complete temporal features. DenseNet and BLSTM were further combined to construct the DenseNet BLSTM model. This model can adaptively extract multi-level features from signals. BLSTM is used to predict results from both the front and back directions to obtain more complete feature information. Simultaneously, it can maximize the utilization of feature information extracted from each layer of network.

IV. THE APPLICATION EFFECT ANALYSIS OF IMAGE GENERATION AND RECOGNITION

This chapter analyzes the application effect of the proposed method in visual image feature recognition of mobile robots.

TABLE I. TEST RESULTS OF SIX METHODS IN BERKELEY IMAGE SEGMENTATION DATASET/%

Methods	F1-measure	Precision	Recall	RI	ACC
NormTree	56.28±13.22	45.17±15.36	72.0±17.06	62.62±8.44	58.75±11.74
LOG	47.37±14.27	35.28±16.26	81.28±4.43	42.54±11.36	46.58±15.32
Ncuts	51.76±12.57	72.12±10.04	48.47±21.07	72.73±14.07	54.57±12.59
Otsu	54.39±13.22	46.14±16.72	81.51±14.44	61.59±7.85	55.34±11.32
K-means	47.14±8.54	74.43±13.15	41.39±7.56	72.11±11.43	45.88±8.67
Ours	63.45±10.12	61.57±12.86	85.82±11.53	74.85±11.33	61.47±12.58

This includes improving the visual image segmentation performance of the K-means and the recognition performance of the DenseNet BLSTM model.

A. Visual Image Segmentation Effect

Firstly, the proposed K-means algorithm incorporating image spatial structure constraints is validated for the visual image segmentation effect of mobile robots. The dataset used in the experiment is the Berkeley image segmentation dataset, which includes benchmark codes, real human annotations, and 500 natural images. The proposed method is compared with classic Otsu algorithm, Laplacian of Gaussian (LOG), Normalized Cuts (Ncuts), NormTree, and traditional K-means algorithm. Five different evaluation indicators are used for quantitative evaluation of image segmentation effectiveness, i.e., F1 measure, accuracy, precision, recall, and Rand index (RI). All images are repeated 10 times in the Berkeley image dataset. Table I illustrates the results.

From Table I, in the comparison of F1 values, the NormTree algorithm is (56.28±13.22) %. The proposed K-means algorithm is (63.45±10.12) %, which is significantly superior to the other five methods. In the comparison of Recall, RI, and ACC, the proposed methods were (85.82±11.53) %, (74.85±11.33) %, and (61.47±12.58) %, respectively. All are the best of the five methods, indicating that they can effectively balance accuracy and recall. The comparison results obtained by all methods on different datasets are significantly different. This could stem from notable disparities amid the data samples within the two datasets, as well as the differing collection methods for the data in each dataset. This method has better performance. To further verify the performance, experiments are conducted again by changing the hyperparameter α . The F1-measures of the proposed method and K-means under different α conditions are shown in Fig. 7.

From Fig. 7, as the hyperparameter α increases, the F1 value of the traditional K-means does not change and remains stable at 0.55. For the proposed K-means, when α is 0, the F1 value is 0.55. As the value of α increases, the F1 value of this method also increases. When the value of α is 10, F1 reaches a maximum of 0.83. Then F1 descends. When the value of α is 20, the F1 value is 0.72, which is still significantly better than the K-means algorithm. The changes in ACC and RI indicators under different hyperparameters α are illustrated in Fig. 8.

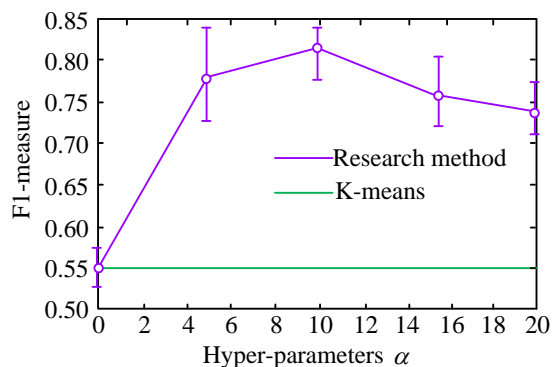


Fig. 7. The impact of different α value on F1 measure and its comparison with K-means clustering method.

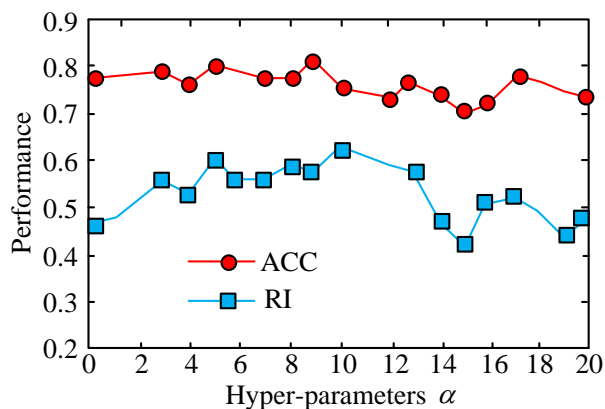


Fig. 8. The impact of different α values on accuracy (ACC) and Randall Index (RI).

From Fig. 8, with the continuous increase of hyperparameter α , the ACC and RI indicators of the proposed method have fluctuated to varying degrees. Among them, the ACC indicator curve has relatively small changes. Most of them are stable between 0.7 and 0.8. When the hyperparameter α values are 9 and 15, the maximum and minimum values of the ACC index appear, which are 0.82 and 0.69, respectively. In terms of RI indicators, when the hyperparameter α is less than 10, the overall trend shows an upward trend, reaching a maximum of 0.65. When the α exceeds 10, the overall RI index shows a downward trend. However, the minimum is maintained above 0.4, and the performance is still relatively good.

B. Analysis of Visual Image Feature Recognition Results

After verifying the proposed image segmentation method, the image feature recognition performance of the constructed DenseNet BLSTM model is further analyzed. The ImageNet dataset is selected for experiments. This dataset is a large visualization database used for visual object recognition software research, which contains more than 20000 categories. It has a large number of pictures. 200 images are randomly selected from four batches for image feature recognition, denoted as groups A, B, C, and D. The proposed DenseNet-BLSTM model is compared with four methods, namely, CNN, RNN, CRNN, and DenseNet. The obtained results are shown in Fig. 9.

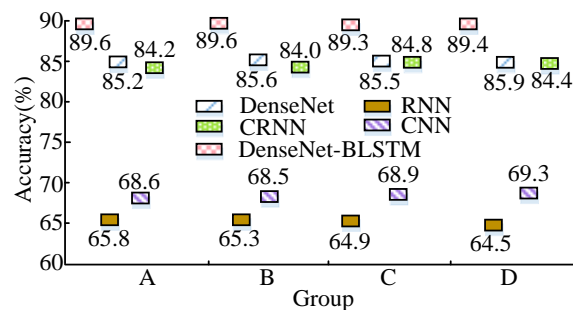


Fig. 9. Image feature recognition results of five models in ImageNet dataset.

From Fig. 9, in the four selected image feature recognition tests, the accuracy of the RNN model is 65.5%, 65.3%, 64.9%, and 64.5%, respectively, concentrated around 65%. The feature recognition accuracy of CNN is 68.6%, 68.5%, 68.9%, and 69.3%, all around 68%. The accuracy of CRNN and DenseNet models is relatively similar. The two fluctuate around 84% and 85% respectively. The four test results of the proposed DenseNet-BLSTM model are 89.6%, 89.6%, 89.3%, and 89.4%, all approaching 90%. Compared with the other four methods, this model has high accuracy and significant performance advantages. Further experiments are conducted on the efficiency of image feature recognition. The Pascalvoc2012 and Cityscapes datasets are used for testing. Among them, the proportion of training and testing sets is 70% and 30%. The recognition time changes of the five methods in the two selected datasets are shown in Fig. 10.

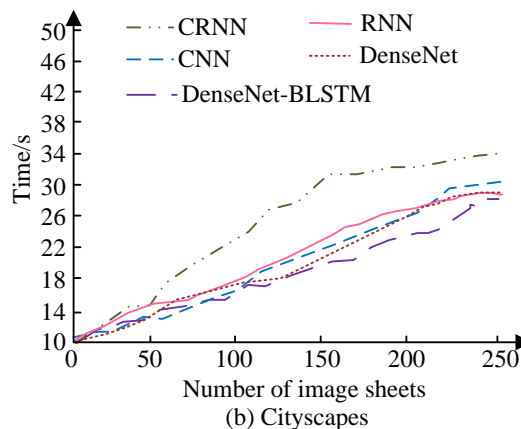
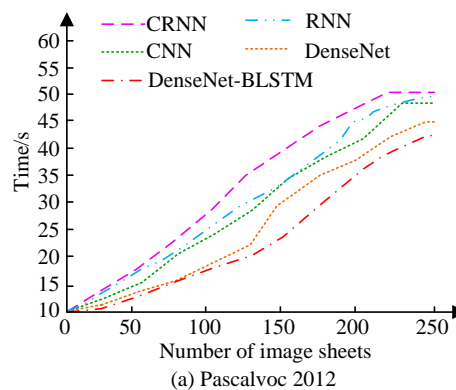


Fig. 10. Five methods for identifying time changes in the two selected datasets.

From Fig. 10(a), in the dataset Pascalvoc2012, as the images increases, the recognition time of all five models increases. Among them, CRNN takes the longest time. When the number of images reaches 250, the time consumption increases to around 50s. The time difference between CNN, RNN, and DenseNet models is relatively small, but they are better than the CRNN model. Among them, the DenseNet model takes the highest time of around 45s. The proposed DenseNet BLSTM model has a maximum duration of 40s, which is the shortest among the five models. From Fig. 10(b), in the dataset Cityscapes, CRNN takes up to 34s, which is the longest among the five methods. The proposed method still takes the shortest time, with a maximum of only 28s. The feature recognition efficiency is high and the performance advantage is significant. Finally, the improved K-means clustering combined with the DenseNet-BLSTM model is applied to the visual image feature recognition of mobile robots. To enhance the persuasiveness of the results, five main scenarios are selected for experiments, namely, mining, risk management, agriculture, factories, and construction. The image feature recognition results before and after the combinations are compared, as shown in Fig. 11.

From Fig. 11(a) in the image feature recognition of mining, risk management, agriculture, factories, and construction scenes, the combination of the previous method achieves the highest recognition accuracy of 79.3% and the lowest score of 70.4% occurs in agricultural and construction settings. Meanwhile, the lowest error is 9.9%, and the highest value is 45.6%, indicating poor classification performance. From Fig. 11(b), after the combination of the proposed method, the accuracy of image feature recognition for buildings is the highest, reaching 96.1%. The lowest value appears in mining scenarios, at 94.8%. Compared with the recognition results before the combination, the combined method has significant performance advantages in visual image feature recognition of mobile robots. The accuracy is between 94.8% and 96.1%. The application effect is better. The results above suggest that the study's method, consisting of CNN and clustering algorithms, is better suited for unsupervised learning tasks, large datasets, and image processing for data clustering.

V. DISCUSSION

With the emergence of artificial intelligence and big data technology, an increasing amount of unlabeled data has become available. It is crucial to utilize the information within the data to explore its potential value. Unsupervised learning employs clustering algorithms to effectively address these challenges, with the K-means clustering algorithm being a classic example. Additionally, in real-world scenarios, data can possess unique structural constraints. Merely applying the K-means clustering algorithm to solve problems without accounting for the data's inherent features frequently results in suboptimal outcomes. Therefore, in response to this issue, we combine the K-means clustering algorithm with the structural characteristics of the data itself and introduce a combination of CNN and K-means clustering to develop a method for recognizing visual images in mobile robots. Throughout the entire experimental process, the study first utilized K-means to process spatial information and introduced image spatial structure constraints for visual image segmentation. Next, a densely connected network is added to the CNN, and combined with a bidirectional long short-term memory network to achieve recognition and segmentation of visual image features. The proposed models in the research have clear and objective functions, all solved using the maximum expectation algorithm. The effectiveness of the algorithm has been verified through a large number of experiments. The K-means clustering algorithm considers spatial constraints in images by integrating positional relationship information between adjacent pixels into the algorithm. This creates an overall framework with a unified objective function. The K-means clustering algorithm with hierarchical constraints achieves hierarchical clustering by establishing a hierarchical tree that can be globally iterated and updated, thereby better mining the hierarchical structure of the data itself. In addition, in complex and dynamically changing scenarios, deep learning methods introduce convolutional neural network models which demonstrate higher accuracy and robustness.

However, due to the continuous changes in environmental factors, regularly updating models has become the key to ensuring the long-term and efficient operation of mobile robots. Online learning and incremental learning technologies provide

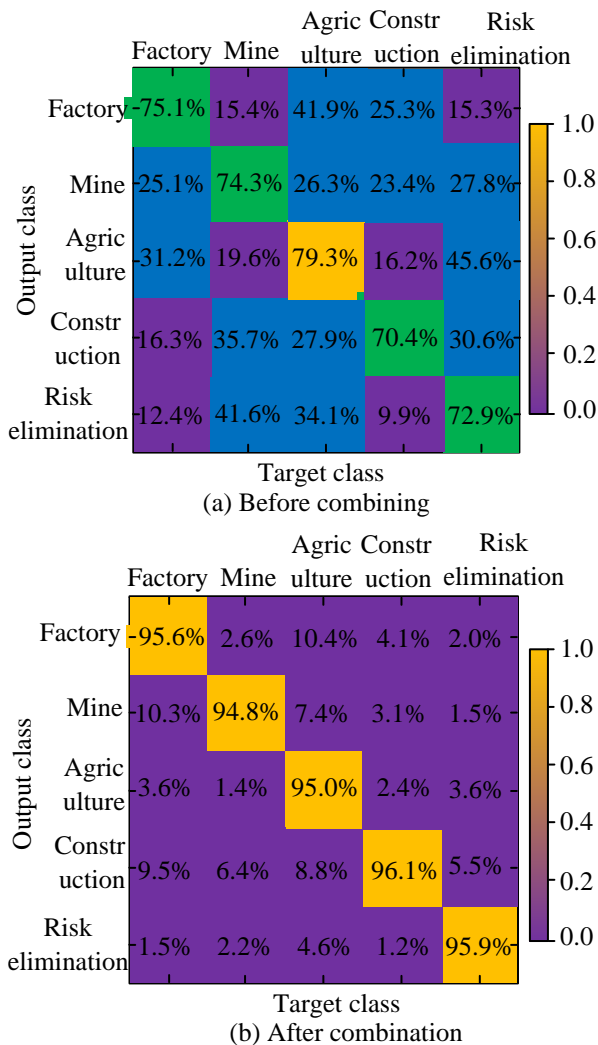


Fig. 11. The recognition results of image features before and after the combination of the proposed method.

effective methods for this. Overall, machine vision based image feature recognition for mobile robots is a versatile challenge that involves algorithm selection, computational efficiency, environmental factors, real-time requirements, and model training and updating. With the advancement of technology, this study aims to develop more efficient and robust methods in the future to meet the practical application needs of mobile robots in various environments.

VI. CONCLUSION

The improvement of visual image processing technology is an important foundation for the wider application of mobile robots. Firstly, an improved K-means algorithm using image spatial structure constraints is proposed. This method is applied to visual image segmentation. The increase in the layers in the CNN network results in complex parameters. Therefore, the DenseNet structural model is introduced and combined with BLSTM to achieve visual image feature extraction. According to the findings, in the comparison of F1 values, the NormTree algorithm is (56.28 ± 13.22) %. The proposed K-means algorithm is (63.45 ± 10.12) %, which is significantly superior to the other five methods. As the hyperparameter α increases, the F1 value of the traditional K-means algorithm stabilizes at 0.55. The proposed K-means algorithm has an F1 value of 0.55 when α is 0. As the value of α increases, the value of F1 also increases. It reaches the maximum of 0.83 when the α is 10. In terms of RI indicators, when the α is less than 10, the overall trend shows an upward trend, reaching a maximum of 0.65. When the α value of the hyperparameter exceeds 10, the overall RI index shows a downward trend. However, the minimum is above 0.4. The performance is still relatively ideal. The image feature recognition performance of the DenseNet BLSTM model constructed is analyzed. In the ImageNet dataset, the four test results are 89.6%, 89.6%, 89.3%, and 89.4%, all approaching 90%. In the dataset Pascalvoc2012, the maximum time consumption of this model is only 40s. In the image feature recognition of five scenes including mining, risk elimination, agriculture, factory and building, the accuracy of the recognition model combined with the improved K-means is between 94.8% and 96.1%, with a high accuracy. However, the proposed method has poor real-time performance. In the future, dimensionality reduction technology is needed to reduce computational complexity. The recognition speed of the algorithm will be further improved.

REFERENCES

- [1] P. Rosenberger, A. Cosgun, R. Newbury, J. Kwan., V. Ortenzi, P. Corke, and M. Grafinger, "Object-independent human-to-robot handovers using real time robotic vision", *IEEE Robot. Autom. Lett.*, vol. 6, pp. 17-23, January 2021.
- [2] Y. Guo, Z. Mustafaoglu, and D. Koundal, "Spam detection using bidirectional transformers and machine learning classifier algorithms", *J. Comput. Cogn. Eng.*, vol. 2, pp. 5-9, April 2023.
- [3] L. Jiang, W. Nie, J. Zhu, X. Gao, and B. Lei, "Lightweight object detection network model suitable for indoor mobile robots", *J. Mech. Sci. Technol.*, vol. 36, pp. 907-920, February 2022.
- [4] H. Kim, H. Kim, S. Lee, and H. Lee, "Autonomous exploration in a cluttered environment for a mobile robot with 2D-Map segmentation and object detection", *IEEE Robot. Autom. Lett.*, vol. 7, pp. 6343-6350, July 2022.
- [5] J. Zan, "Research on robot path perception and optimization technology based on whale optimization algorithm", *J. Comput. Cogn. Eng.*, vol. 1, no. 4, pp. 201-208, March 2022.
- [6] J. Jiang, Y. Bie, J. Li, X. Yang, G. Ma, Y. Lu, and C. Zhang, "Fault diagnosis of the bushing infrared images based on mask R-CNN and improved PCNN joint algorithm", *High Voltage*, vol. 6, pp. 116-124, December 2021.
- [7] K. C. Chen, Y. W. Huang, G. M. Liu, J. W. Liang, Y. C. Yang, and Y. H. Liao, "A hierarchical k-means-assisted scenario-aware reconfigurable convolutional neural network", *IEEE Trans. Very Large Scale Integrat. (VLSI) Syst.*, vol. 29, pp. 176-188, January 2021.
- [8] B. Ragupathy and M. Karunakaran, "A deep learning model integrating convolution neural network and multiple kernel K means clustering for segmenting brain tumor in magnetic resonance images", *Int. J. Imaging Syst. Technol.*, vol. 31, pp. 118-127, September 2021.
- [9] D. J. I. Z. Chen, "Automatic vehicle license plate detection using K-means clustering algorithm and CNN", *J. Electrical Eng. Autom.*, vol. 3, pp. 15-23, March 2021.
- [10] Z. Rustam, S. Hartini, R. Y. Pratama, R. E. Yunus, and R. Hidayat, "Analysis of architecture combining convolutional neural network (CNN) and kernel K-means clustering for lung cancer diagnosis", *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 10, pp. 1200-1206, June 2020.
- [11] Y. Liu, Y. Yixuan, and M. Liu, "Ground-aware monocular 3D object detection for autonomous driving", *IEEE Robot. Autom. Lett.*, vol. 6, pp. 919-926, April 2021.
- [12] A. Sungheetha and R. Sharma, "3D image processing using machine learning based input processing for man-machine interaction", *J. Innov. Image Process. (JIIP)*, vol. 3, pp. 1-6, February 2021.
- [13] X. Wang, L. T. Yang, L. Song, H. Wang, L. Ren, and M. J. Deen, "A tensor-based multiattributes visual feature recognition method for industrial intelligence", *IEEE Trans. Ind. Inform.*, vol. 17, pp. 2231-2241, March 2021.
- [14] I. J. Jacob and P. E. Darney, "Design of deep learning algorithm for IoT application by image based recognition", *J. ISMAC*, vol. 3, pp. 276-290, September 2021.
- [15] S. Niu, B. Li, X. Wang, and H. Lin, "Defect image sample generation with GAN for improving defect recognition", *IEEE Trans. Autom. Sci. Eng.*, vol. 17, pp. 1611-1622, July 2020.
- [16] S. Oslund, C. Washington, A. So, T. Chen, and H. Ji, "Multiview robust adversarial stickers for arbitrary objects in the physical world", *J. Comput. Cogn. Eng.*, vol. 1, pp. 152-158, September 2022.
- [17] L. Wüthrl, C. Pylatiuk, M. Giersch, F. Lapp, T. von Rintelen, M. Balke, and R. Meier, "DiversityScanner: robotic handling of small invertebrates with machine learning methods", *Mol. Ecol. Resour.*, vol. 22, pp. 1626-1638, May 2022.
- [18] Y. Wang, Y. Liu, W. Feng, and S. Zeng, "Waste haven transfer and poverty-environment trap: evidence from EU", *Green Low-Carbon Econ.*, vol. 1, pp. 41-49, February 2023.
- [19] L. C. Ngugi, M. Abelwahab, and M. Abo-Zahhad, "Recent advances in image processing techniques for automated leaf pest and disease recognition-A review", *Inform. Process. Agr.*, vol. 8, pp. 27-51, February 2021.

Explore Chinese Energy Commodity Prices in Financial Markets using Machine Learning

Yu Cui¹, Tianhao Ma²

HITSZ School of Economics and Management, Harbin University of Technology (Shenzhen), Shenzhen, 518055, China¹
College of Business and Economics, Australian National University, Acton, 2601, Australia²

Abstract—This study simultaneously investigates the causality and dynamic links between international energy trade and economic price changes, especially in the Chinese commodity market. To get a causal route, it attempts to identify the linear and nonlinear causality among commodity prices, equities, and the exchange rate in China and the United States (US). Here, we adapt multilayer perceptron networks to obtain a nonlinear autoregressive model for causality discovery. After comparing methods without networks, this study proves that the nonlinear causality discovery method using machine learning performs best on simulated data. Subsequently, we apply that causality to actual data; we combine the causal routes, particularly from the machine learning methodology, to investigate the existence of a causal direct or indirect relationship among Chinese commodity prices, long-term interest rates, stock index, and exchange rates in China and the US. The steady-state accuracy of cmlpgranger is 99%. In most cases, the order of judgment accuracy of causality is cmlpgranger > HSICLasso > ARD > LinSVR. The results show that Energy trade as an element of the global economic system. The Chinese commodity price of energy has an interactive relationship with the Chinese commodity price of agricultural products. The significant transmission is from the commodity price of energy to equities, then to the exchange rate, and, finally, to the commodity price of agricultural products.

Keywords—Chinese commodity price; exchange rate; stock markets; machine learning; international energy trade; global economic system

I. INTRODUCTION

As an essential part of the global industrial chain, the price fluctuation of commodities has a tremendous impact on the real economy, and the inflation level has a significant effect. The supply and demand pattern of commodities has changed dramatically since 2020. The COVID-19 pandemic raised energy costs, political instability and increased demand for energy [1], which led to intense fluctuations in international commodity prices. In the post-pandemic era, the evolution of the world pandemic situation is out of sync with the economic recovery of various countries. Under the pressure of green transformation and global development, major developed economies have implemented large-scale fiscal stimulus and super-wide monetary policies. At the same time, there have been profound changes in the international economic and trade landscape, geopolitical factors, extreme climate, and other factors. Global supply and demand patterns. Commodities are generally in short supply [2]. Under the influence of various complex factors, it is very important to sort out the causal logic among the factors affecting commodity prices and clarify the impact of commodity price changes on the macro economy for effec-

tively adjusting policies and overcoming the global economic recession.

China is currently the world's largest consumer country [3]. The continued demand for energy, essential raw materials, and agricultural products has rapidly increased China's commodity imports. China has become the world's largest importer of commodities such as iron ore, aluminum ore, lead ore, nickel-chromium ore, and crude oil [4]. The existing literature in 2018 includes discussions on factors that affect commodity prices. However, the current global indicators mainly consider the role of developed countries, but do not consider China's role and its impact on China.

This study also aims to provide a comprehensive analysis using the machine learning (ML) approach. Although numerous studies have analyzed the factors and effects of international commodities prices, we make full use of the algorithm advantages of ML to integrate our findings with those of previous studies. The majority of applied analyses on this topic used the vector autoregressive (VAR) Granger's tests. Sahlian et.al. used the Granger test to establish the causality between the market capitalization and financial indicators [5]. In the classic linear VAR method, when evaluating the cause of Granger, the maximum delay must be stated. When the relationship between the past of one series and the future of another series does not belong to the model category. The method based on model in the real world may fail [6-8]. This usually occurs when there is a non-linear relationship between the past and future of a series. The nonlinear relationship between the past and the future can be detected by minimizing assumptions about the predicted relationship [9]. For example: Yin Z et al. used multiple repetitive neural networks to demonstrate the effectiveness of nonlinear random time series models. [10] Neural networks can display complex nonlinear and non-auxiliary interactions between inputs and outputs. Some studies introduced the structural learning framework in multi-layer sensor (MLP) and Recurrent neural network (RNN). This led to Granger's nonlinear causal discovery. However, the use of these methods to analyze causal processes between multiple variables is particularly rare in the field of economics.

The surge in oil and food prices over the past decade has prompted a large amount of research to focus on the common flow of crude oil and agricultural products. Then, some studies begin modeling to investigate correlations and connections or pathways. On the contrary, some studies provide evidence of the neutral relationship structure between crude oil and agricultural products. Based on the research of previous genera-

tions, the significance and innovation of this study can be summarized in the following three points.

Firstly, we are the first to adopt the ML method instead of traditional economic models. The determination of causal relationships in existing literature is mainly based on the assumptions of econometric models. However, most of these models explore linear relationships. Based on our research, we are the first to use ML to study the causal discovery and relationship between energy and agricultural futures prices. Using ML in this study, we can break the assumptions of the Traditional economy model and find more nonlinear relationships. In addition, compared to other traditional econometric models that only estimate one or two parameters, the ML time size Helping people study time series in time zones and frequency domains

Secondly, we have important evidence in studying the relationship between energy prices and agricultural product prices. In past research, there has been no consensus on the research conclusions, price indicators, and research methods for energy and agricultural product prices at different periods. Country/Region. This study focuses on the prices of China's energy and agricultural futures markets, providing many research conclusions for the literature.

Finally, this study discussed whether this relationship exists and attempted to express causal relationships based on important financial indicators. Most research on the impact of commodity prices focuses on two pairs of relationships. However, this study consists of two or more factors of causality diagram, thus obtained a comprehensive causal path. The media focused on Chinese stock market prices and the exchange rate between China and the United States.

The structure of the remaining parts of this article is as follows: In Section I, we introduced early research related to commodity market reasons or correlations. In Section II, we use motivational datasets to evaluate the effectiveness of machine learning causal discovery methods. After determining the cause-finding ability of these methods, we will use it to analyze the actual data and get the results in Section III. Finally, Section IV summarizes the article.

II. METHODOLOGY AND DATA

A. Adapting Neural Networks for Granger Causality

The Nonlinear Autoregressive Model (NAR) allows XT to dynamically evolve based on typical nonlinearity [11]:

$$x_t = g(x_{<t1}, \dots, x_{<tp}) + e_t \quad (1)$$

where $x_{<ti} = (\dots, x_{(t-2)i}, x_{(t-1)i})$ represents the past of sequence i; we assume that the additive noise zero mean e_t .

In a forecasting setting, the mutual modeling of nonlinear functions g typically uses neural networks. Neural networks have a long history in predicting NAR using traditional architectures [12] and the latest deep learning techniques [15]. These methods use MLP, where the input

$$x_{<t} = x_{(t-1):(t-K)} \quad (2)$$

Our main approach is to model each output g_i using a separate MLP to easily clarify the impact of input on output. We call it component-wise MLP (cMLP) [17]. Set g_i in the form of MLP with L-1 layers, and let the vector $h_t^l \in \mathbb{R}^H$ denote the values of the m-dimensional lth hidden layer at time t. The discovery and research of nonlinear causal relationships have become more widespread, for example, M Roso et al (2022) not only focused on theory, but also on how to build Python packages to complete testing [18].

Definition 1. Time series J is a non-causal relationship of time series I if all $(x_{<t1}, \dots, x_{<tp})$ and all $x'_{<tj} \neq x_{<tj}$,

$$g_i(x_{<t1}, \dots, x_{<tj}, \dots, x_{<tp}) = g_i(x_{<t1}, \dots, x'_{<tj}, \dots, x_{<tp}) \quad (3)$$

that is, g_i is invariant to $x_{<tj}$.

The parameters of the neural network are given by weights W and biases b as each layer, $w = \{w^1, \dots, w^L\}$ and $b = \{b^1, \dots, b^L\}$. To compare with the time series VAR model, we divided the weight of the first layer into time delays, $W^1 = \{W^{11}, \dots, W^{1K}\}$. The parameter sizes include $W^1 \in \mathbb{R}^{H \times pK}$, $W^l \in \mathbb{R}^{H \times H}$ for $1 < l < L$, $W^L \in \mathbb{R}^H$, $b^1 \in \mathbb{R}^H$ for $1 < L$ and $b^L \in \mathbb{R}$. Using this notation, the vector of first layer hidden values at time t is given by

$$h_t^1 = \sigma \left(\sum_{k=1}^K W^{1k} x_{t-k} + b^1 \right) \quad (4)$$

where σ is an activation function [19].

In Equation (4), if the jth column of the first layer weight matrix, W_j^{1k} , contains zeros for all k, then series j does not Granger-cause series i. That is, $x_{(t-k)j}$ for all k does not affect the hidden cell h_t^1 . So according to the definition 1 output x_{ti} , we can see that g_i divided by $x_{<tj}$ remains unchanged.

B. Creating Benchmark Datasets

1) *Group1: Lorenz-96 model*: We use these two methods to detect Granger's causal network from P's simulated Lorenz-96 data. Find the impact of many attributes in different ways [11].

The continuous dynamics of the P Vilorenz model will be obtained from the following styles.

$$\frac{dx_{ti}}{dt} = \left(x_{t(i+1)} - x_{t(i-2)} \right) x_{t(i-1)} - x_{ti} + F \quad (5)$$

where $x_{t(-1)} = x_{t(p-1)}$, $x_{t0} = x_{tp}$, $x_{t(p+1)} = x_{t1}$ and F is a mandatory constant that determines the degree of nonlinearity and chaos in the set.

2) *Group2*: Nonstationary data: Due to the fact that most of the attributes we want to check are not smooth in practice, we will add trends to the Lorenz-96 P dimension data to obtain the data. The second set of extended Dickey Fuller (AD-Fuller) checks can be used to test the unit root directory within a single variable process in the presence of sequence relationships. We use ADFuller to test whether the data is stable.

$$\text{Trend}_t = \text{Trend}_{t-1} + (\text{Trend}_1 - \text{Trend}_0) / I \quad (6)$$

Here, $t \in (0, I)$, I is the length of features and Trend is linear.

As shown in Table I, we take P as 5 and obtain the results of two datasets with P values:

TABLE I. ADFULLER TEST OF DATA (CALCULATED BY AUTHORS)

Group1	Adfuller P_Value		Group2	Adfuller P_Value	
X_1	6.641034e-22	stationary	$X_1 + \text{Trend}$	0.689549	Nonstationary
X_2	5.266443e-18	stationary	$X_2 + \text{Trend}$	0.716672	Nonstationary
X_3	5.218544e-10	stationary	$X_3 + \text{Trend}$	0.700451	Nonstationary
X_4	3.249791e-14	stationary	$X_4 + \text{Trend}$	0.706434	Nonstationary
X_5	8.472980e-12	stationary	$X_5 + \text{Trend}$	0.797429	Nonstationary

C. Comparing Models for Granger Causality

We compare four methods on different numbers of features using stationary and nonstationary data. The four methods are LinSVR2, automatic relevance determination (ARD), HSI-CLasso, and CMLP-Granger.

The basic support vector regression (SVR) concepts are introduced by Schölkopf and Smola; Linear SVR (LinSVR) is a special case of SVR with a linear axis. Huang and Cai improved SVR and used these methods to select features. The advantage of LinSVR is that due to the small complexity of the model, the

results are easy to interpret. However, it cannot recognize nonlinear relationships that typically occur in real data.

ARD was proposed by MacKay based on the Bayes model, which effectively selects relevant features through preliminary training. Evaluate initial hyperparameters by maximizing the probability in the data. This process is called evidence augmentation, or maximizing the second type of likelihood. Wipf and Nagarajan [20] applied the ARD method to prune large numbers of irrelevant features, leading to a sparse explanatory subset and demonstrating that the ARD prior maintains advantages compared to conventional priors in feature selection.

TABLE II. RESULTS OF DIFFERENT METHODS (CALCULATED BY AUTHORS.)

Methods	x_num	Stationary	Nonstationary
		Accuracy	Accuracy
ARD	5	100%	60%
HSICLasso	5	100%	60%
LinSVR	5	76%	60%
cmlpgranger	5	80%	72%
ARD	15	56%	76%
HSICLasso	15	92%	67%
LinSVR	15	80%	80%
cmlpgranger	15	100%	97%
ARD	35	63%	64%
HSICLasso	35	96%	85%
LinSVR	35	63%	63%
cmlpgranger	35	99%	98%
ARD	50	59%	59%
HSICLasso	50	98%	90%
LinSVR	50	58%	58%
cmlpgranger	50	99%	97%
ARD	100	56%	55%
HSICLasso	100	98%	85%
LinSVR	100	54%	54%
cmlpgranger	100	98%	97%

Freidling et al. [21] improved the least absolute shrinkage and selection operator (Lasso) method on feature selection by integrating Lasso and particular kernel functions with kernel-based independence measures such as the Hilbert–Schmidt independence criterion (HSIC). Compared to the former Lasso methods, the HSICLasso can capture nonlinear dependency with a clear statistical interpretation and deal with high-dimensional feature selection scenarios in which the number of training samples is smaller than that of features.

In 2021, Alex Tank, Ian Covert et al. [22] use time series normal neural network Model selection framework of Granger nonlinear causality. These researchers applied the multi-layer sensor module (cMLP) and the long run and short run memory architecture of the module. (cLSTM) includes related sparsity, which promotes penalties for network inbound weight, thus selecting Granger demonstrated the construction of Granger causality diagram, which is the correct basis for linear and nonlinear settings.

In Table II, the results can be summarized into three points. (1) In most cases, the order of judgment accuracy of causality is $cmlpgranger > HSICLasso > ARD > LinSVR$. (2) The performance of $cmlpgranger$ varies with the number of factors. When the number of distinguishing factors is greater than 5, it is far better than other methods. (3) $cmlpgranger$ also performs well for nonstationary data. When the data are nonstationary, the accuracy of the method in judging causality is significantly higher than that of other models.

III. RESULTS

Through simulated data, proved $cmlpgranger$ in nonlinear nonstationary has good capability of causal relationships found in the data. Therefore, we determine the cause and effect of the actual data of the Chinese futures market by integrating the results of several causality methods. Specifically, we want to ascertain how the relationship between the commodity prices of energy affects that of agricultural products and its specific impact path.

In order to assess the causal relationship of commodity prices, we have adopted from the Ministry of Commerce of China's commodity price index (CCPI) international commodity prices. Financial indicators covering the stock market, bond market, futures market, interest rate, credit market, and foreign exchange market are extracted from the Wind Economic Database as financial market factors affecting commodities prices. All data begin from June 2006 to October 2021 monthly. The nomenclature of indicators is shown in Table III.

Finally, for each method, we obtain a causality matrix, where $a_{i,j}$ in the matrix equal to 1 indicates that the j th factor is the cause of the i th factor. Otherwise, there is no apparent causality between the two elements. To acquire accurate results, we synthesize the conclusions of the models that perform well on the simulation dataset. Table IV shows the causality result confirmed by $cmlpgranger$. Table V shows the causality result proved through the HSICLasso and ARD methods.

TABLE III. NOMENCLATURE (CALCULATED BY AUTHORS.)

Full Name	Abbreviation	Full Name	Abbreviation
China commodity price index_ Total index	CCPI_T	S&P 500 index	SP500
China commodity price index_ Energy	CCPI_E	CSI 500 Index	CSI500
China commodity price index_ Non-ferrous metals	CCPI_M	US discount rate	US_DR
China commodity price index_ Agricultural products	CCPI_A	CN discount rate	CN_DR
The US dollar/RMB exchange rate	USDCNY	US10-year Treasury yield	US_TB10Y
The Dollar index	DI	CN10-year Treasury yield	CN_TB10Y

TABLE IV. CAUSALITY RESULTS OF THE CMLPGRANGER METHOD (CALCULATED BY AUTHORS.)

Cmlpgranger	CCPI_E	CCPI_M	CCPI_A	US_TB10Y	USDCNY	DI	CSI500	SP500	CN_TB10Y
CCPI_E	1	1	0	0	1	0	0	0	0
CCPI_M	0	1	0	0	0	0	0	1	0
CCPI_A	0	0	0	0	0	0	0	1	0
US_TB10Y	0	1	0	0	0	0	0	0	0
USDCNY	0	0	0	0	0	0	0	0	0
DI	0	0	0	0	1	1	0	0	0
CSI500	1	0	0	0	1	0	1	1	0
SP500	0	1	0	0	0	0	0	1	0
CN_TB10Y	0	1	0	0	0	0	0	0	1

TABLE V. INTERSECTION OF CAUSALITY RESULTS OF THE HSICLASSO AND ARD METHODS (CALCULATED BY AUTHORS.)

HSICLasso&ARD	CCPI_E	CCPI_M	CCPI_A	US_TB10Y	USDCNY	DI	CSI500	SP500	CN_TB10Y
CCPI_E	1	1	0	0	0	1	0	0	0
CCPI_M	0	1	1	0	1	1	0	0	0
CCPI_A	0	0	1	0	1	1	0	0	0
US_TB10Y	0	1	1	1	0	0	0	0	0
USDCNY	0	0	1	0	1	1	0	0	1
DI	0	0	0	0	0	1	1	0	1
CSI500	1	0	0	0	0	0	1	0	0
SP500	0	0	0	0	0	0	0	0	0
CN_TB10Y	0	0	0	0	0	0	0	0	0

Combining the results of the two methods, we can use Fig. 1 to show the relationship between futures market prices and other economic indicators. The bold black line shows the causal relationship, which is proved by three methods (cmlpgranger, HSICLasso, and ARD). The black line indicates the causal relationship, which is proved by two methods (HSICLasso and ARD). The red dotted line shows the causal relationship proved by cmlpgranger method, which can be indirectly verified by the other two methods. The arrow points from cause to result.

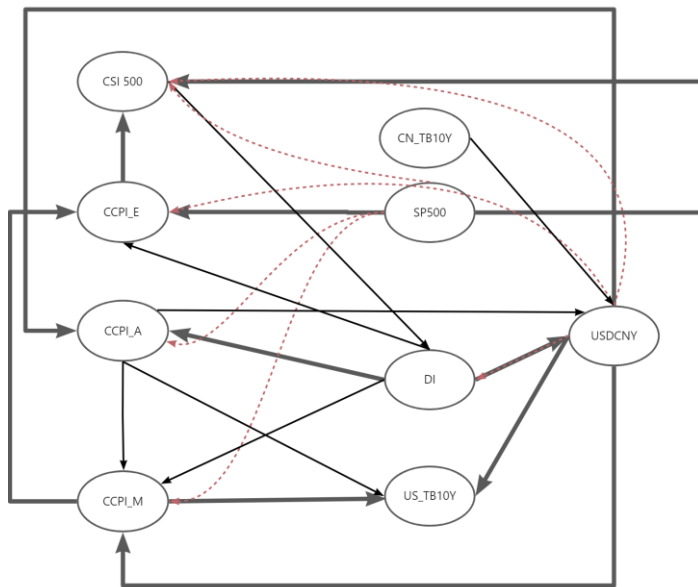


Fig. 1. Causality routes (Calculated by authors.).

Thus, we can conclude that the CCPI_E and CCPI_A have an interactive relationship. The Chinese commodity price of energy will affect the CSI500, thereby affecting the dollar index. The similar conclusion given by Shabir et. al. show that the impact of oil prices and exchange rate on stock prices exists and varies across bullish, bearish, and normal states of the stock market [23]. [23] The dollar index will also affect the exchange rate between China and the US and eventually transmit to the China commodity price of agricultural products. Conversely, the China commodity price of agricultural products will affect the China commodity price of non-ferrous metals and finally transmit to energy prices.

Other researchers have also investigated the conclusion that CCPI_E affects the direction of CCPI_A, and analyzed the relationship between oil prices. Food (agricultural product prices) and exchange rates, as oil prices are transmitted to the local agricultural market through exchange rates. Nazlioglu and Soytaş conducted a study on global oil and agricultural prices, explaining the relative strength of the US dollar [24]. Awartani et al. pointed out that the oil market is the main source and risk transfer for stocks, euro/dollar exchange rates, precious metals, and agricultural products [25]. In our results, we observe the evidence favoring this argument with respect to Chinese commodity prices.

IV. CONCLUSIONS

In this study, we use monthly data from the Chinese financial market from June 2006 to October 2021 to investigate the causal route between the commodity prices of energy and agricultural products. Unlike traditional econometric models, we apply the ML method for the first time to examine this issue. Using these methods, we find that the energy commodity price is vital. It can affect equities changes represented by the CSI 500 Index, thus affecting the exchange rate between China and the US and eventually causing changes in the prices of agricultural products. In comparison with previous studies, we find a correlation between the energy commodity price and agricultural products but emphasize indirect media in causality.

The conclusions drawn in this study led to further considerations. First, we should remain alert to changes in energy commodity prices. Since the commodity price of energy is likely to affect the Chinese capital market price. Second, the exchange rate influences the commodity prices of different varieties in China, indicating China is still highly dependent on foreign countries in terms of the import of products involved in the commodity market. In 2022, under the general environment of long-term inflation in the US, the exchange rate between China and the United States breaking 7, the stalemate in the Russia-Ukraine war, and the continued slowdown in global economic growth, the price of China's futures market will also be volatile under the influence of the strong US dollar. Third, based on the results, most causal paths reflect that the prices of various commodities in China or the prices of China's capital markets are always affected. Still, there is no obvious path to show the impact of China's factors on American financial

indicators. This suggests that there is still a long way to go to improve the Chinese commodity market.

But there is still much work to be done. First, causality may change over time. While we consider the beginning year of COVID-19, the actual time when significant economic changes will occur is 2-3 years after COVID-19, not 2021. As new data becomes available, we should pay attention to these changes. Second, more research is needed to prove the effectiveness of machine learning methods in doing causal discovery.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Fig. 1. Causality routes; Table I: ADFuller test of data; Table II: Results of different methods; Table III: Nomenclature; Table IV: Causality results of the cmlpgranger method; Table V: Intersection of causality results of the HSICLasso and ARD methods.

REFERENCES

- [1] Smal, T., Wieprow, J. Energy Security in the Context of Global Energy Crisis: Economic and Financial Conditions. *Energies* 2023, 16, 1605. <https://doi.org/10.3390/en16041605>
- [2] Yun, L., Cui, X. M., Xiao, L. S., et al. International Commodity supply and demand analysis framework: Global perspective and China's role. *International Economic Review*, 2022(03):68-88+5.
- [3] Lin, J., Fridley, D., Lu, H., et al. Has coal use peaked in China: Near-term trends in China's coal consumption. *Energy Policy*, 2018, 123(DEC.):208-214.
- [4] Saiful, Alim, Rosyadi, et al. Impact of Donald Trump's tariff increase against Chinese imports on global economy: Global Trade Analysis Project (GTAP) model. *Journal of Chinese economics and business studies*, 2018.
- [5] Sahlian, D.N., Popa, A.F., Nicoară, Ș.A., Bâtcă-Dumitru, C.G. Examining the Causality between Integrated Reporting and Stock Market Capitalization. The Case of the European Renewable Energy Equipment and Services Industry. *Energies* 2023, 16, 1398. <https://doi.org/10.3390/en16031398>
- [6] T. Terasvirta, D. Tjøstheim, C.W. J. Granger " et al., *Modelling Nonlinear Economic Time Series*. Oxford University Press Oxford, 2010.
- [7] H. Tong, "Nonlinear time series analysis," *International Encyclopedia of Statistical Science*, pp. 955–958, 2011.
- [8] B. Lusch, P. D. Maia, and J.N. Kutz, "Inferring connectivity in networked dynamical systems: Challenges using Granger causality," *Physical Review E*, vol. 94, no. 3, p. 032220, 2016.
- [9] P.-O. Amblard and O.J. Michel, "On directed information theory and Granger causality graphs," *Journal of Computational Neuroscience*, vol. 30, no. 1, pp. 7–16, 2011.
- [10] Yin, Z., Barucca, P. Deep Recurrent Modelling of Granger Causality with Latent Confounding. 2022.
- [11] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily Fox. Neural granger causality for nonlinear time series. arXiv preprint arXiv:1802.05842, 2018.
- [12] S. A. Billings, *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. John Wiley & Sons, 2013.
- [13] S. R. Chu, R. Shoureshi, and M. Tenorio, "Neural networks for system identification," *IEEE Control Systems Magazine*, vol. 10, no. 3, pp. 31–35, 1990.
- [14] S. Billings and S. Chen, "The determination of multivariable nonlinear models for dynamic systems using neural networks," 1996.
- [15] Y. Tao, L. Ma, W. Zhang, J. Liu, W. Liu, and Q. Du, "Hierarchical attention-based recurrent highway networks for time series prediction," arXiv preprint arXiv:1806.00685, 2018.
- [16] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Graph convolutional recurrent neural network: Data-driven traffic forecasting," arXiv preprint arXiv:1707.01926, 2017.
- [17] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Multistep neural networks for data-driven discovery of nonlinear dynamical systems," arXiv preprint arXiv:1801.01236, 2018.
- [18] Roso, M., Myńczak, M., Cybulski, G. Granger causality test with nonlinear neural-network-based methods: Python package and simulation study. *Computer methods and programs in biomedicine*, 2022, 216:106669.
- [19] Joris M Mooij, Jonas Peters, Dominik Janzing, Jakob Zscheischler, and Bernhard Schölkopf. Distinguishing cause from effect using observational data: methods and benchmarks. *Journal of Machine Learning Research*, 17(32):1–102, 2016.
- [20] Wipf, D. P., Nagarajan, S. S. A New View of Automatic Relevance Determination. International Conference on Neural Information Processing Systems. Curran Associates Inc. 2007.
- [21] Freidling, T., Poignard, B., Climente-González, H., et al. Post-selection inference with HSIC-Lasso. International Conference on Machine Learning. PMLR, 2021.
- [22] Tank, A., et al. (2021) "Neural Granger Causality. " *IEEE Transactions on Pattern Analysis and Machine Intelligence* 01.
- [23] Hashmi, S. M., Chang B H, Huang L, et al. Revisiting the relationship between oil prices, exchange rate, and stock prices: An application of quantile ARDL model. *Resources Policy*, 2022, 75: 102543.
- [24] Nazlioglu, S., Soytaş, U. Oil price, agricultural commodity prices, and the dollar: A panel cointegration and causality analysis. *Energy Economics*, 2012, 34(4): 1098-1104.
- [25] Awartani, B., Aktham, M., & Cherif, G. (2016). The connectedness between crude oil and financial markets: Evidence from implied volatility indices. *Journal of Commodity Markets*, 4(1), 56–69.

Research on the Application of Multi-Objective Algorithm Based on Tag Eigenvalues in e-Commerce Supply Chain Forecasting

Man Huang, Jie Lian*

College of Business Administration, Fujian Business University, Fuzhou 350016, China

Abstract—With the continuous development of Internet technology, the scale of Internet data is increasing day by day, and business forecasting has become more and more important in corporate business decision-making. Therefore, to improve the accuracy of Multi Target Regression in the actual e-commerce supply chain forecasting, research through the method of constructing the labeling feature for each target is optimized, the Multi-Target Regression via Sparse Integration and Label-Specific Features algorithm is obtained, and the experimental analysis is carried out on the performance of the algorithm and the application effect in the actual e-commerce supply chain. The experimental results show that the average of Relative Root Mean Square Error value of the research algorithm and is the lowest in most datasets, with a minimum of 0.058 in the effect experiments of prediction and label-specific features; in the effect and flexibility experiments of sparse sets, the lowest average of Relative Root Mean Square Error value of the research algorithm was 0.058, and the average rank value was the smallest. In addition, the average of Relative Root Mean Square Error value of the research algorithm is the smallest under the target variable of Y2 in the Enb data, and its value is 0.075. In the actual e-commerce supply chain forecast, the research algorithm has the highest score of 0.097 points. Overall, research algorithm has a better forecasting effect and higher performance, and has better practicality in practice, and can play a better effect in actual e-commerce supply chain forecasting.

Keywords—Label features; multi-objective algorithm; sparse set; e-commerce supply chain; multi target regression

I. INTRODUCTION

In the business decision-making of enterprises, planning and control are very critical, and forecasting is the basis for the control planning and forecasting of future trends, and forecasting is also important to avoid one-sided decision-making and mistakes [1]. In traditional regression analysis, the marker and target variables are often single, but a single object variable cannot accurately describe the complex information contained. Therefore, the single target regression analysis method has been unable to predict objective problems well and accurately, and Multi-Target Learning (MT) came into being [2]. At the same time, MT-based multi-target regression (Multi Target Regression, MTR) has also been gradually paid attention to, which refers to the use of a common set of input variables to predict multiple continuous variables. In this regard, a large number of domestic and foreign scholars have carried out a lot of research. Based on MTR, Nabati et al. proposed a Gaussian regression-related

algorithm to avoid the fitting problem in training [3]. Based on multi-objective regression, Osojnik et al. proposed a contour tree to improve the prediction effect of multi-label classification [4]. Syed et al. conducted research on semi-supervised techniques based on multi-objective regression by analyzing limited instances, thereby achieving efficient prediction of new objects [5]. However, current multi-objective regression methods mainly focus on mining the correlation between targets and handling the complex relationship between input and output, and most methods learn models from the same feature space, resulting in insufficient flexibility and low prediction performance. At the same time, there is not much research on its application in e-commerce supply chain prediction. Based on this, the study obtained a multi-objective regression via Sparse integration and Label Specific Features (SI-LSF) algorithm by improving MTR, aiming to effectively improve the flexibility of processing multi-objective datasets and thereby enhance the accuracy of the algorithm in e-commerce supply chain prediction.

The research is divided into six sections. Section I is the Introduction of the study. The Section II is a summary and discussion of the current research on multi-objective optimization methods. Section III is the study of the SI-LSF algorithm in e-commerce supply chain prediction, including the definition and related methods of multi-objective regression, and the optimization of the SI-LSF algorithm for multi-objective regression problems. Section IV analyzes the performance and practical application of the SI-LSF algorithm. Section V presents the discussion and last Section VI is a summary of the entire article.

II. RELATED WORK

With the rapid development of information technology, the amount of data on the network is increasing day by day, and the expression of data is also changing rapidly, which makes the ability of data analysis and data mining more and more important. At the same time, the ability to predict data is required. It is also becoming more and more important, and MT has gradually gained attention as a method that is more in line with the laws and characteristics of objective things. Based on this, scholars at home and abroad have carried out research on it. By using multi-objective optimization, Das et al. proposed a fast and accurate meta-model to accurately calculate the effective degree of losses and oil spills, reducing the risk at sea [6]. Dong performed multi-objective optimization of hybrid composites based on multi-objective

*Corresponding Author

regression to verify the effectiveness of the positive hybrid effect in improving the flexural strength of materials [7]. Irodov et al. solved the multi-objective optimization problem of pellet burners by performing a multi-objective regression optimization analysis on the work of tubular gas burners [8]. Wang et al. solved the multi-objective constrained optimization problem of variables by introducing random forests and other methods to continuously approach the objective and constraint functions [9]. Based on the tripartite competition mechanism, Han et al. proposed an improved multi-objective particle swarm optimization algorithm to effectively solve the problems of multi-objective diversity and poor convergence performance [10]. Pereira et al. developed a numerical model of complex structures by proposing equi-grid multi-objective optimization with six objectives, thereby reducing the instability of such grid tubes [11]. Based on reinforcement learning, Dang et al. proposed a multi-objective optimized resource allocation model to achieve a better-distributed search [12]. Ullah et al. achieved multi-objective optimization of motor switching by proposing a new permanent magnet forward pole with flux bridges etc [13].

In addition, Grover et al. comprehensively study e-commerce and supply chain management, to purchase raw materials according to demand forecast, which greatly facilitates supply chain management [14]. Li et al. established a digital model based on signal game theory to obtain revenue forecast information, and then proposed the optimal information acquisition strategy in the supply chain [15]. Yang et al. identified the importance of forecast updates in supply chains by studying the pricing problem in a two-tier fashion supply chain [16]. Shen et al. determined the driving effect of shared information on supply chain management by comprehensively analyzing the problem of forecasting information sharing in supply chain management so that it can better match supply chain demand [17]. Li et al. built an effective forecasting model to reasonably control the multi-item inventory in the supply chain [18]. Chaudhuri et al. integrated extreme learning machines to propose an optimized forecasting model, thereby realizing real-time accurate forecasting of products in supply chain management [19]. Wan realizes the risk prediction of the supply chain by proposing a risk prediction model related to the manufacturing industry and ensures the healthy development of enterprises [20]. Proposing a combined model, Jaipuria et al. constructed a hybrid forecasting technology for supply chain demand, thereby ensuring inventory safety and sufficient order quantity in the replenishment cycle of goods [21].

Through the research of domestic and foreign scholars, it can be found that the multi-objective method can predict the future based on summarizing the laws of objective things, and predictive analysis is very important in the e-commerce supply chain. Therefore, it is expected that the proposed SI-LSF algorithm will be helpful in the actual prediction of the e-commerce supply chain.

III. RESEARCH ON E-COMMERCE SUPPLY CHAIN FORECASTING BASED ON THE SL-LSF ALGORITHM

A. Analysis of Correlation Methods based on Multi-Objective Regression

To improve the breadth and accuracy of the MTR method in e-commerce supply chain forecasting, the research proposes SI-LSF and analyzes its performance and application in actual forecasting by creating a special marker feature for each target. MTR refers to the existence of multiple dependent variables in a model, which takes into account the mapping between multiple output objects. These mapping relationships can be linear or nonlinear. In addition, for the MTR problem, a class of samples contains multiple different output variables, and there are often different semantic relationships between them [22]. By mining the correlation of each output object, the comprehensive prediction effect of multiple indicators can be effectively improved. It is assumed here that X represents the input feature space, Y stands for the output target space, and the membership formulas of the two are shown in Eq. (1) and (2).

$$\begin{cases} X \subset R^m \\ X = (X_1, X_2, \dots, X_m) \end{cases} \quad (1)$$

In Eq. (1), R represents the datasets and m represents the number of feature vectors.

$$\begin{cases} Y \subset R^d \\ Y = (Y_1, Y_2, \dots, Y_d) \end{cases} \quad (2)$$

In Eq. (2), d represents the number of target variables. Therefore, the input vector and the output vector can be given in the MTR problem, and the related equations are shown in Eq. (3).

$$\begin{cases} x^{(l)} = (x_1^{(l)}, \dots, x_m^{(l)}) \\ y^{(l)} = (y_1^{(l)}, \dots, y_d^{(l)}) \end{cases} \quad (3)$$

In Eq. (3), $x^{(l)}$ and $y^{(l)}$ represents a sample, which l represents the target variable, and its number range is between $[1, n]$, which n represents the number of samples. On this basis, the relevant equation of the training set can be given as Eq. (4) shown.

$$D = \left\{ (x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)}) \right\} \quad (4)$$

In Eq. (4), D represents the training set. Therefore, it can be determined that the task of MTR is to learn a mapping function, so that for any uncertain vector, all output variables can be obtained at the same time through $\hat{y} = h(x)$. The multi-objective regression problem and the multi-label problem are similar in that the output variable of the multi-objective regression problem is continuous, while the classification problem is discrete, so it can be divided into two types, namely problem transformation method and algorithm adaptation method, the contents of which are shown in Fig. 1.

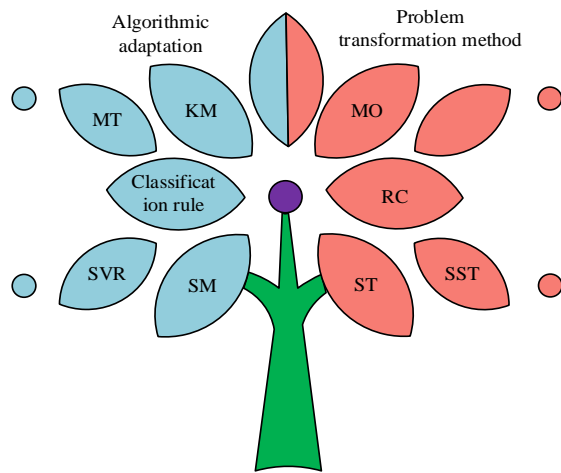


Fig. 1. Specific classification of multi-objective regression methods.

It can be seen in Fig. 1, the problem transformation method mainly includes Single-Target Method (ST), Stacked Single-Target (SST), Regressor Chains (RC), and multi-target support vector regression (Multi-Output Support Vector Regression, MO), algorithm adaptation methods include statistical methods (Statistical Methods, SM), support vector machines (Support Vector Regression, SVR), kernel methods (Kernel Methods, KM), regression tree (Multi-Target) Regression Trees, (MT) and classification rules. The ST problem is often used to deal with the MTR problem, so ST is used for learning prediction. The least squares criterion is used to solve it, and its formula is shown in Eq. (5).

$$\{\hat{a}_{ij}\}_{j=1}^m = \arg \min \{ \hat{a}_{ij} \}_{j=1}^m \left\{ \sum_{i=1}^N \left(y_i^{(l)} - \sum_{j=1}^m a_{ij} x_j^{(l)} \right)^2 \right\} + \lambda_i \sum_{j=1}^m a_{ij}^2 \quad (5)$$

In Eq. (5), \hat{a}_{ij} represents the regression coefficient, i represents the target variable, j represents the feature vector, and $\lambda_i > 0$ represents the ridge parameter. In the MO problem, multiple virtual samples can be constructed by virtualizing the eigenvectors, and the objective function definition formula is shown in Eq. (6).

$$\begin{cases} f = \frac{1}{2} \|w\|^2 + \sum_{l=1}^N \frac{1}{c} \sum_{i=1}^d e_i^{(l)2} \\ s.t. y_i^{(l)} = w^T \phi(I_i, x^{(l)}) + I_i b + e_i^{(l)} \end{cases} \quad (6)$$

In Eq. (6), f represents the objective function, w represents the weight vector, c represents the trade-off factor between dimension and loss, $e_i^{(l)}$ represents the fitting error, $\phi(\cdot)$ represents the nonlinear transformation of the feature space, b represents the deviation vector, I represents the virtual sample, T represents the transpose.

The algorithm adaptation method uses a single model to predict all the targets at one time, so that the correlation and internal connection between the target and the input features can be captured, which has the advantages of more

interpretability and better prediction performance. In the adaptation method, the most important are SM and SVR. SM is believed to treat the first attempt as predicting multiple targets at the same time, and its purpose is to improve the prediction accuracy by using the correlation between target variables, while SVR is usually used to deal with single-target regression problems. It reflects the relationship between input variables and output variables on a certain datasets, and the operation formula is shown in Eq. (7).

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^N L \left(y_i^{(l)} - \left(\phi(x^{(l)})^T w + b \right) \right) \quad (7)$$

In Eq. (7), w represents the regressor, C represents the user-selected parameters, and b represents the bias. In addition, the MT method in the algorithm adaptation method can predict multiple continuous targets when dealing with multi-objective problems. It has two advantages, that is, the scale of a single multi-objective regression tree model is often smaller than that of a single multivariate model. The multi-target regression tree can better distinguish the correlation of multiple target variables; the KM method is an index kernel with a coupled regularization function; the classification rule is to convert the regression tree set into a rule set, to select the best subset of rules to improve prediction accuracy.

B. Research on SI-LSF Algorithm Optimized for Multi-Objective Regression Problem

MTR types are mainly problem transformation methods and algorithm adaptation methods, and their basic functions have been widely used in computer vision and medical image analysis. However, current MTR is mainly studied for linear models but lacks a method capable of simultaneously handling the nonlinear relationship of multiple objects with the input space [23]. Therefore, the research optimizes the existing problems of MTR and proposes the SI-LSF algorithm. The SI-LSF algorithm is based on the superposition of a single target, adopts the method of boosting learning, and studies the specific characteristics of the markers to explore multiple targets. Inter relationships between variables. Second, a sparse ensemble-based model is built using the sparse ensemble method. Finally, combining sparse integration with target features can simultaneously address two challenges of MTR within a single framework, namely establishing the fundamental relationship between input features and output objects and exploring their inter-correlation to improve prediction performance. The stacking of a single target is also called Stacked Single Target (SST) [24]. Its training and prediction framework is shown in Fig 2.

It can be seen in Fig. 2, its frame structure is divided into two stages of training and two stages of prediction. In the training phase, the first is to train the model for each target through the training set, to obtain the predicted value, and then enter the second phase of training, that is, to establish the second phase of training and then output the final training model. In the prediction stage, for an unknown sample, first pass the model in the first stage to obtain the prediction vector, then extend the prediction vector into the feature vector, and then set the result as the new input predicted by the model in

the second stage value, and finally use the second-stage model to obtain the final prediction result. The core idea of SST is to use the predictions of the remaining targets in the first stage as additional input variables. Although the real value of the target can be obtained in the training set, there is no real target in the prediction set. Therefore, in the prediction process, the SST must depend on the estimation of the target based on the prediction results of the single-target regression. However, such an approach violates a central assumption of supervised learning theory: training and trial data must be consistent and independent of each other [25].

In the second stage, due to the introduction of additional input variables, the noise of the prediction is caused, which leads to the estimation error of the model. Therefore, to solve this problem, it is necessary to improve the training and prediction of the target variable. The numerical value used in the prediction compatibility. Label-Specific Features (LSF) is another important element in SI-LSF, and its framework is shown in Fig.3.

It can be seen in Fig. 3, before each object is marked for recognition, it is first preprocessed and extended as an additional feature. The training set transformed by the instruction is $D' = \left\{ \left(x^{(j)} \cup \hat{y}^{(j)}, y^{(j)} \right) \mid 1 \leq j \leq n \right\}$, which $x^{(j)}$ represents the initial feature vector and $y^{(j)}$ is the target real

value vector. In the learning process of LSF, the first step is to find the relevant features for each target, to improve its learning accuracy. The formulas are shown in Eq. (8), (9), and (10).

$$R_1(\gamma, s) = \{X_i | x_{i\gamma} \leq s\}, R_2(\gamma, s) = \{X_i | x_{i\gamma} > s\} \quad (8)$$

In Eq. (8), it x_i represents a certain column attribute, s represents x_i a certain value in it, which is represented as the corresponding split point, R_1 and R_2 represents s the two parts of the attribute using the division, and the value in the γ representation X represents a feature.

$$\hat{c}_t = \frac{\sum_{X_i \in R_1(\gamma, s)} y_{ij}}{|R_1|} \quad (9)$$

In Eq. (9), it \hat{c} represents the corresponding target mean value, t which is a determined value, which is 1 or 2.

$$\min_{\gamma, s} \left(\min_{X_i \in R_1(\gamma, s)} \sum (y_{ij} - \hat{c}_1)^2 \right) + \min_{\gamma, s} \left(\min_{X_i \in R_2(\gamma, s)} \sum (y_{ij} - \hat{c}_2)^2 \right) \quad (10)$$

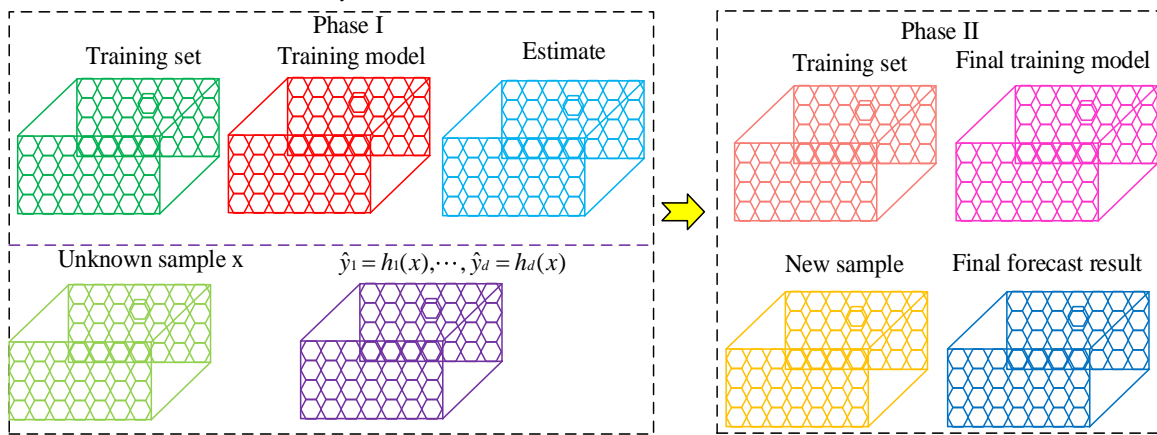


Fig. 2. SST training and testing framework diagram.

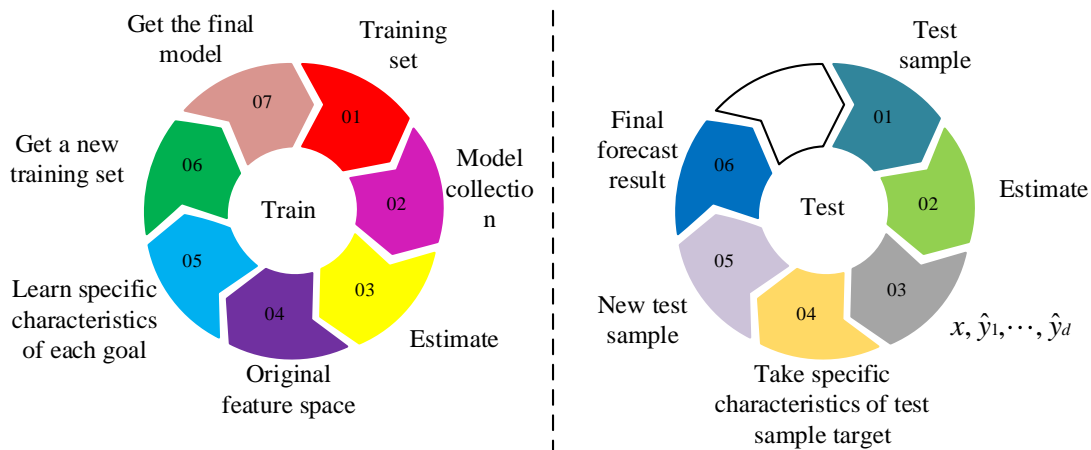


Fig. 3. LSF frame structure diagram.

According to the minimized square error in Eq. (10), the best division feature and split point can be found, to determine the relevant feature set of the target and delete redundant features. It is worth noting that these features are only a subset of features in the training set, and there is not enough diversity in the feature space to represent objects well. To improve the difference of the feature space, the target correlation method is used to construct the marker-specific feature, and the residual of each repetition is used to describe the local feature, and it is used as an additional feature to reflect the local information of the feature space. Perform direct modeling to improve the expressiveness of the model. The relevant calculation formulas are shown in Eq. (11), (12), and (13).

$$g^0(X'_{TR}) = f_q((X'_{TR}); Y_j) \quad (11)$$

Set the initial value represented by Eq. (11). Among them, g represents the predicted value, X'_{TR} represents the output space, f_q represents the basic learner, and Y_j represents the target.

$$r_i^{(j)} = - \left[\frac{\delta l(Y_j, g_{t-1}(X'_{TR}))}{\delta g_{t-1}(X'_{TR})} \right] \quad (12)$$

Eq. (12) means to find the negative value of the current mode and treat it as a residual. Among them, r represents the relevant features and l represents the squared error loss function.

$$g_t(X'_{TR}) = g_{t-1}(X'_{TR}) + f_q(X'_{TR}; r_i^{(j)}) \quad (13)$$

Equation (13) is to set the target as the residual estimation in Eq. (12), to use the negative gradient value to update the model, and use it as the target of the next iteration. In addition, after the introduction of marker features, LSF still selects a single-target regression analysis method, which will cause the complex input and output relationship models in the first and second stages to be unable to be established. Therefore, the proposal of SI-LSF is exactly to flexibly handle these complex relationships. SI-LSF is a sparse ensemble algorithm, which will be applied to ensemble learning. Therefore, the research proposes a corresponding aggregation function to deal with the complex problem of ensemble learning calculation, and its calculation formula is shown in Eq. (14) and (15).

$$w_j^* = \min_{w_j} \frac{1}{2} \sum_{i=1}^N (w_j^T \hat{y}_i - y_{ij})^2 + \lambda \|w_j\|_1 \quad (14)$$

In Eq. (14), w_j represents the weight vector and $\lambda \|w_j\|_1$ represents the regular term. It means that by introducing a regular term, the weight of the base model becomes 0, which realizes the sparseness of the data and automatically selects a regression method suitable for learning, which improves the flexibility of the algorithm. And reduce the time and space complexity and enhance the prediction effect.

$$h_j = \sum_{i=1}^k w_{ij}^* f_i \quad (15)$$

In Eq. (15), h represents the final prediction result, f represents the regression method, and w_{ij} represents the distribution weight. Based on this, the frame structure of SI-LSF can be given, as shown in Fig. 4.

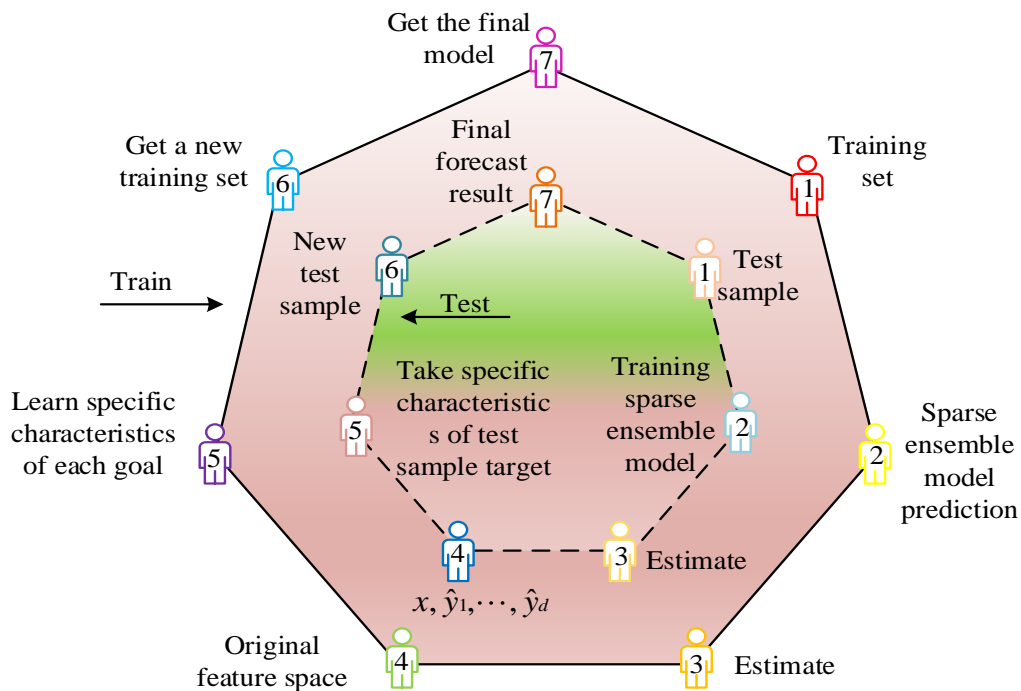


Fig. 4. SI-LSF frame structure diagram.

It can be seen in Fig. 4, unlike most existing methods that integrate the prediction results into the final stage, the SI-LSF algorithm integrates the prediction results of each model, and its advantage is that it can accurately determine the first, the second stage marks, the eigenvalues and selects the most important mode by weighting, thereby reducing the interference of the system and improving the accuracy of the system. Therefore, the framework structure of the SI-LSF algorithm utilizes a sparse ensemble to generate a sparse ensemble model and perform relevant training predictions on it to obtain the specific characteristics of the sample target. Add the target and specific features to form a new training set and test set to get the final model and prediction results.

IV. SI-LSF ALGORITHM PERFORMANCE AND PRACTICAL APPLICATION ANALYSIS

To test the performance of the SI-LSF algorithm, related experiments were carried out. Before the experiment, the study collected related datasets. Since there are relatively few related public datasets, the study selected 18 commonly used datasets for the experiment. The dataset information is shown in Fig. 5.

In Fig. 5, the horizontal axis 1-18 represent 18 data sets, namely Aandro (prediction of future values of water quality variables), Slump (concrete slump), Edm (electronic discharge machining), Atp7d (airline ticket prices), Sfl (solar flares), Oes97 (occupational employment survey), Atpld, Jura (determination of 359 seven heavy metals in soils in the Swiss Jura region), Oes10, Osales (“online product sales”) preprocessed version in competition), Enb (energy building),

Wq (14 target attributes of water quality), Sf2, Scpf (preprocessed version of the dataset used in a competition), S cm20d (supply chain management), R f1 (river flow), R f2, and S cmls. In addition, the three sub-graphs represent the number of samples, the number of features, and the number of targets in the dataset, respectively.

Among them, these datasets are predictions for the future or related content, which can effectively improve the actual response speed of the research algorithm, which is more in line with the overall research. They contain different target attributes in different fields, so they have high effectiveness and practicality in promoting intelligent analysis and improving actual work efficiency. Therefore, this study selected these 18 publicly available multi-objective regression datasets to validate the research algorithm.

On this basis, the research uses the average relative root mean square error (aRRMSE) measure as the evaluation index, and the smaller the value, the better the performance. And introduce the integrated regression chain (Ensemble of Regression Chains, ERC), support vector regression chain (SVR-correlation Chains, SVRCC) and multi-layer multi-target regression (Multi-layer Multi-target Regression, MMR) and SST algorithm, it is combined with the performance comparison of SI-LSF algorithm mainly includes three aspects: prediction effect, the effectiveness of label-specific features, and effectiveness of sparse ensemble. The first is to compare the prediction effects of the experiments. The prediction effects of several algorithms under different datasets are shown in Fig. 6.

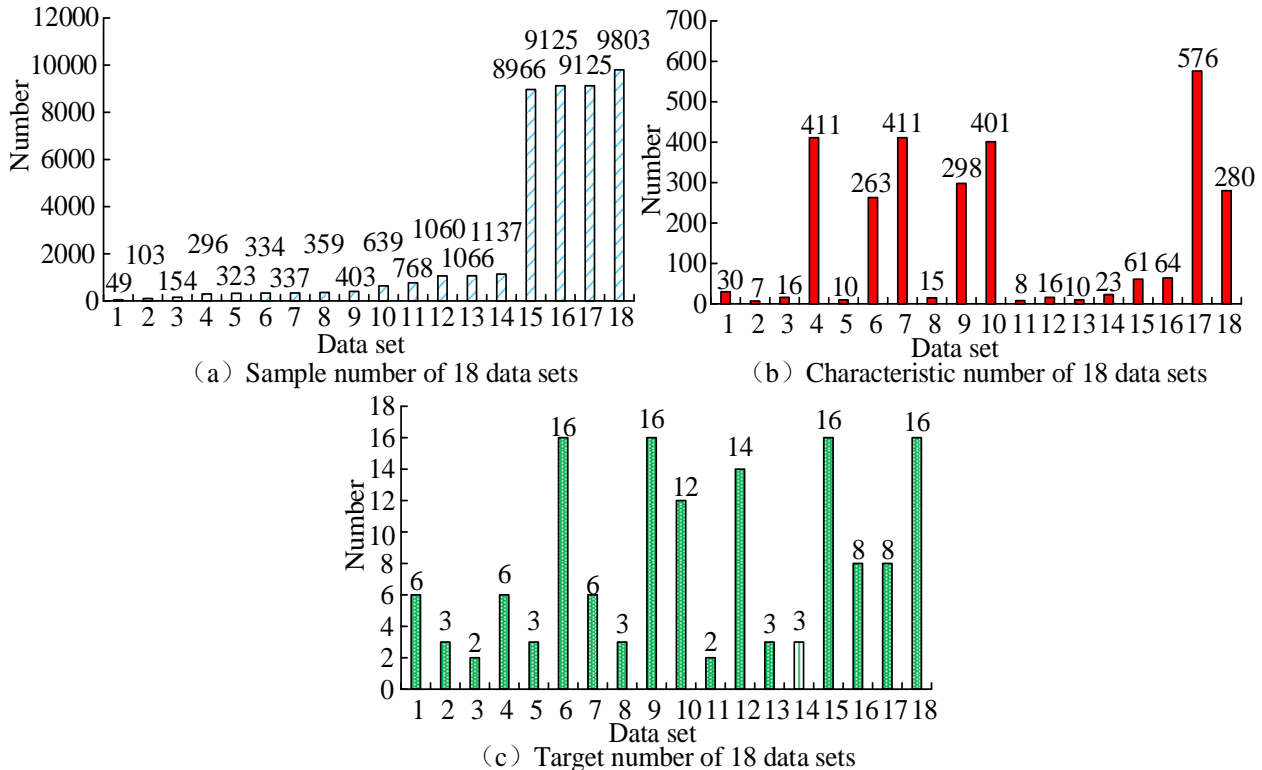


Fig. 5. Relevant information of 18 datasets.

It can be seen in Fig. 6, the SI-LSF algorithm has the lowest aRRMSE values in all 16 datasets, and only the second lowest aRRMSE values in two datasets, roughly between 0 and 0.9. Among them, the SI-LSF algorithm has the smallest Rf2 in the datasets, which is 0.058, and the overall average value is 0.477. In addition, in the average ranking of the five algorithms, SI-LSF also has the lowest aRRMSE value of 1.111. The experimental results show that the prediction effect of the SI-LSF algorithm is the best among the five algorithms, and the performance is also the best. The second goal is to verify the validity of the specific features of the SI-LSF labels. Based on this, the research sets the control scalar as the label-specific feature, compares the SI-LSF algorithm with the sparse integration (Sparse Integration, SI) algorithm, and also compares the aRRMSE values of the two. The results are shown in Fig. 7.

It can be seen in Fig. 7 that the aRRMSE value of the SI-LSF algorithm is smaller than that of the SI algorithm on the 16 datasets, greater than that of the SI algorithm on the Rf1 dataset, and equal to the SI algorithm on the Oes97 dataset. In addition, the aRRMSE value of SI-LSF is generally between 0 and 0.9, and its minimum value appears on the Enb dataset at 0.070. On the whole, the prediction performance of SI-LSF on most datasets has the rain SI algorithm. Through the verification of the specific characteristics of the label, it is proved that it can further improve the SI-LSF algorithm and significantly improve it. The prediction accuracy of the SI-LSF algorithm, so label-specific features are effective for the SI-LSF algorithm to improve. Finally, to verify the effectiveness of the sparse ensemble, the study introduced

Support Vector Regression and Label-Specific Features (SVR-LSF), Linear regression (Linear regression and Label-Specific Features, Linear-LSF), and random forest (The three basic regression models of Random Forest and Label-Specific Features, RF-LSF) are compared with the SI-LSF algorithm. It is worth noting that the SI in the SI-LSF algorithm is used to solve the complex relationship between the input characteristics and the output objects, so it is necessary to use the SI as a control variable to conduct experiments, and the results are shown in Fig. 8.

It can be seen in Fig. 8, in the 15 datasets, the aRRMSE value of SI-LSF is lower than that of the other three algorithms. It is the same as the RF-LSF algorithm in the Atp7d and Wq datasets, which is 0.428. -LSF is the same, which is 0.796; the lowest value of the SI-LSF algorithm appears in the datasets Rf2, which is 0.058, and its average value is 0.477, which is lower than 0.615 of SVR-LSF, 0.695 of Linear-LSF and 0.542 of RF-LSF, its average rank value is one, which is much lower than the other three algorithms. On the whole, the prediction effect of the SI-LSF algorithm is better, and it has the best performance among the four algorithms, which shows that the sparse ensemble is effective in improving the performance of the SI-LSF algorithm. In three aspects, namely the prediction effect, the effectiveness of label-specific features, and the effectiveness of sparse ensemble, it is proved that the SI-LSF algorithm has better performance. In the complex relationship between objects, the study selected seven datasets from 18 datasets and compared the value of aRRMSE for each object, and the results are shown in Table I.

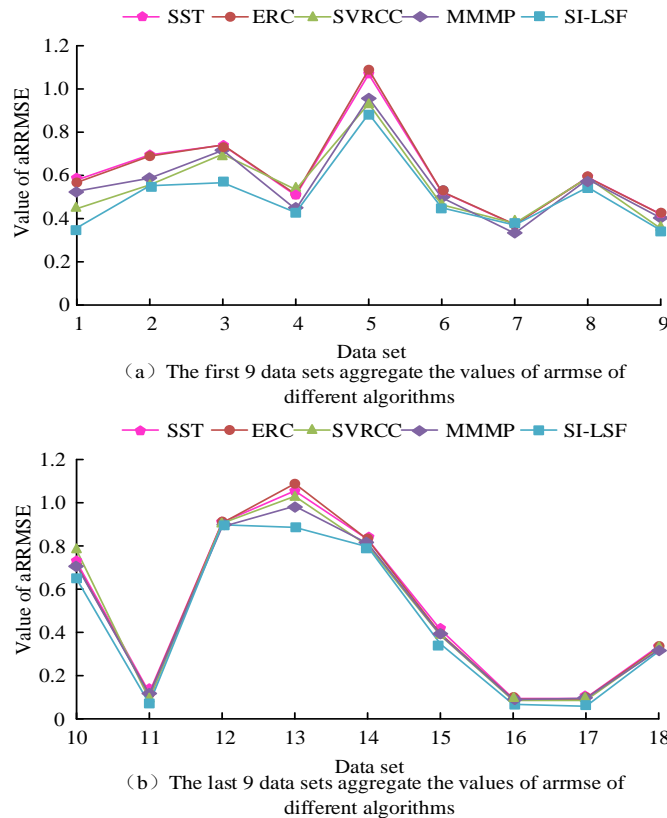
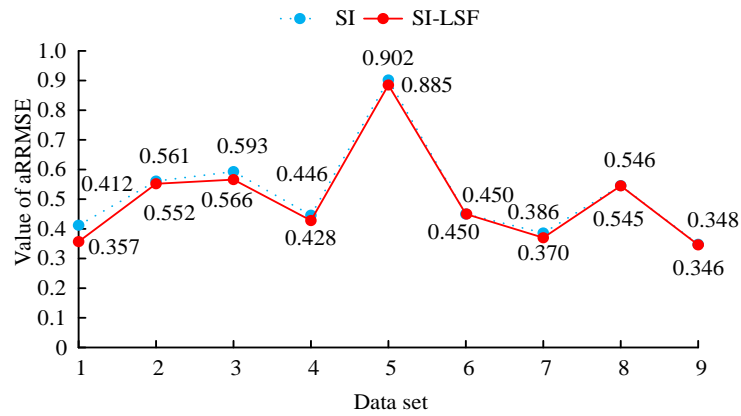
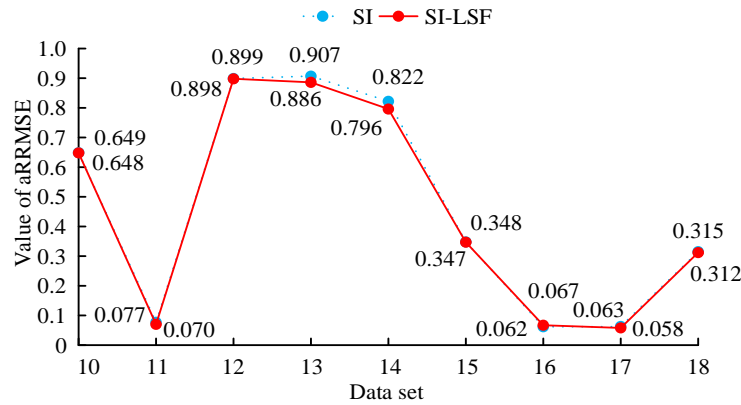


Fig. 6. aRRMSE values of several algorithms under different datasets.



(a) aRRMSE values in the first 9 datasets of the two algorithms

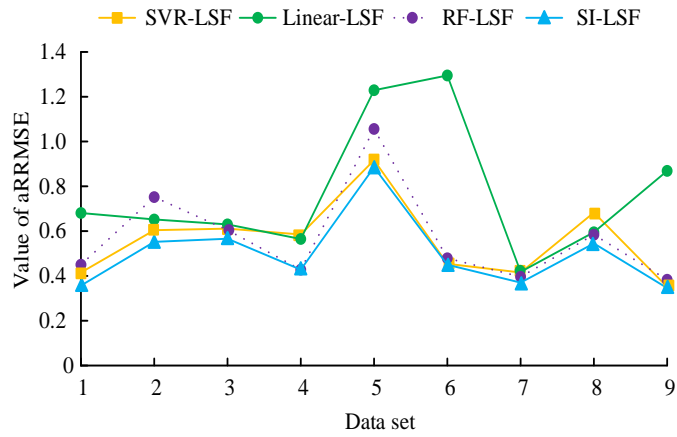


(b) aRRMSE values in 9 data sets after two algorithms

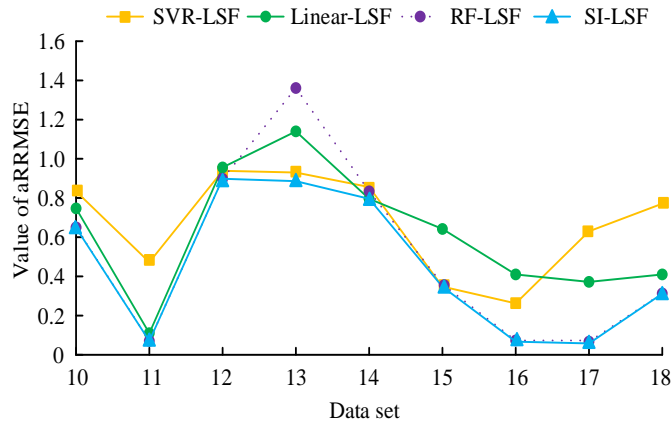
Fig. 7. aRRMSE values in different datasets of the two algorithms.

TABLE I. A RRMSE VALUES OF DIFFERENT TARGET VARIABLES OF FOUR ALGORITHMS IN SEVEN DATASETS

Datasets	Target variable name	SVR-LSF	Linear-LSF	RF-LSF	SI-LSF
S lump	C_S_Mpa	0.155	0.391	0.628	0.155
	FLOW_cm	0.821	0.755	0.756	0.755
	SLUMP_cm	0.822	0.789	0.789	0.781
E dm	DFlow	0.497	0.568	0.557	0.497
	DGap	0.721	0.668	0.642	0.629
S fl	c-class	0.951	0.951	0.951	0.951
	m-class	1.011	0.915	0.873	0.873
	x-class	0.781	1.752	1.215	0.779
Jura _	Cd	0.675	0.700	0.701	0.661
	Co	0.617	0.599	0.538	0.498
	Cu	0.764	0.500	0.541	0.481
Enb _	Y1	0.524	0.075	0.069	0.069
	Y2	0.412	1.213	0.075	0.075
S f2	c-class	1.101	0.955	0.955	0.955
	m-class	0.991	1.101	0.99 0	0.99 0
	x-class	0.751	1.415	2.102	0.751
S cpf	comments	1.044	0.892	0.895	0.887
	views	0.845	0.762	0.775	0.761
	votes	0.789	0.775	0.979	0.732

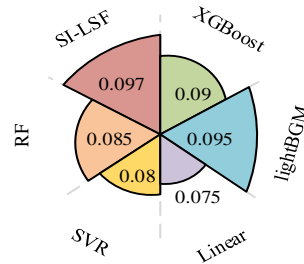


(a) aRRMSE values of the four algorithms in the first nine datasets

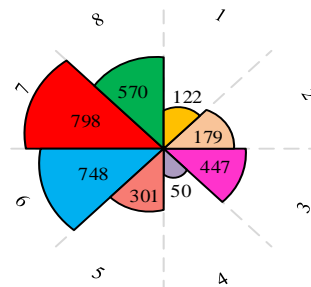


(b) aRRMSE values of the four algorithms in the last nine datasets

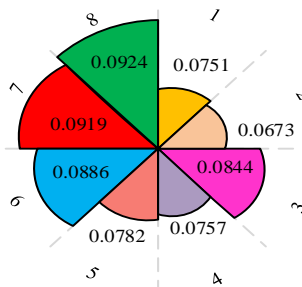
Fig. 8. aRRMSE values of four algorithms in different datasets.



(a) Score value of several algorithms in supply chain demand forecasting



(b) Feature dimension of si-lsf algorithm under different feature groups



(c) Score of si-lsf algorithm under different feature groups

Fig. 9. Comparison score results of several algorithms and score results of SI-LSF.

It can be seen from Table I that the aRRMSE value of SI-LSF on most datasets is the smallest, and in the Enb data, the Y2 target variable is at the smallest value, which is 0.075. In general, in each dataset, there is a complex relationship between the input target and the output feature. At the same time, the SI-LSF algorithm has a sparse integration, so its prediction performance is significantly better than other algorithms, which also proves that the SI-LSF algorithm is effective in processing. It has strong flexibility in multi-objective problems. To further verify the effectiveness of the SI-LSF algorithm in practical applications, the research applies it to the field of the e-commerce supply chains to predict multi-objective tasks for supply chain demand, and it mainly conducts comparative experiments under different feature groups. Here again, two regression models are introduced, namely extreme gradient boosting (eXtreme Gradient Boosting, XGBoost) and distributed gradient boosting framework (light Gradient Boosting Machine, lightGBM). The experimental results are shown in Fig. 9.

In Fig. 9, numbers 1-8 represent feature groups, representing basic statistical features, discrete features, time-series-related features, optimal combination features, basic statistical features + discrete features, basic statistical features + discrete features + time-series-related features, and basic statistics Features + discrete features + time series related features + optimal combination features and feature selection. It can be seen in Fig. 9(a), the scores of several algorithms are roughly between 0.08 and 0.09, of which the score of the SI-LSF algorithm is 0.097, which is the highest among several algorithms, showing that the SI-LSF algorithm is used in the supply chain. The forecasting performance in demand is the highest and has strong applicability. It can be seen in Fig. 9(b) and 9(c), the SI-LSF algorithm has the highest score when the number of feature dimensions is 570, that is, 0.0924. Therefore, it can be determined that the SI-LSF algorithm has the highest score after data preprocessing and feature selection. It can achieve a good forecast in supply chain demand and has strong practicability.

V. DISCUSSION

The SI-LSF algorithm has the lowest aRRMSE value in all 16 datasets, with only the second lowest aRRMSE value in both datasets, roughly between 0 and 0.9. Among them, the SI-LSF algorithm is the smallest in the dataset Rf2, with a value of 0.058 and an overall average of 0.477. In addition, among the average rankings of the five algorithms, the aRRMSE value of SI-LSF is also the lowest, at 1.111. In addition, the overall aRRMSE value of SI-LSF is between 0 and 0.9, with its minimum value appearing on the Enb dataset at 0.070. When comparing with other algorithms, the aRRMSE value of SI-LSF is lower than that of the other three algorithms. In the Atp7d and Wq datasets, it is the same as the RF-LSF algorithm with a value of 0.428, while in the Scpf dataset, it is the same as the Linear-LSF algorithm with a value of 0.796. This result is superior to the results of Wang et al [26]. In practical applications, the SI-LSF algorithm has a score of 0.097, which is the highest among several algorithms. This indicates that the SI-LSF algorithm has the highest predictive performance in supply chain demand and has strong applicability. This result is basically consistent with the results

of Moghadam et al [27]. Overall, the SI-LSF algorithm can achieve good prediction in supply chain demand after data preprocessing and feature selection, and has strong practicality.

VI. CONCLUSION

The advent of the big data era requires enterprises to improve their data mining capabilities, and at the same time requires enterprises to make effective predictions based on these data. Therefore, the research proposes the SI-LSF algorithm by creating special marking characteristics for each target and conducts experimental analysis on its performance and practical application, so as to improve the accuracy of MTR in e-commerce supply chain prediction. The experimental results show that in the prediction effect experiment, the aRRMSE value of the SI-LSF algorithm is the lowest in most datasets, with a minimum of 0.058 and an average of 0.477; in the validity experiment of label-specific features, the aRRMSE value of the SI-LSF algorithm is lower than that of the SI algorithm, and its lowest value is 0.070; in the effectiveness experiment of sparse integration, the aRRMSE value of the SI-LSF algorithm is lower than the other three algorithms in most datasets, and the lowest value is 0.058, and its average ranking value is 1, which is also the lowest; in the flexibility experiment of the SI-LSF algorithm, it is found that the aRRMSE value of the SI-LSF algorithm in most datasets is the smallest, and it is the smallest value in the Y2 target variable in the Enb data, which is 0.075. In addition, in the actual e-commerce supply chain prediction experiment, it is found that the SI-LSF algorithm has the highest score of 0.097. Under different feature groups, the SI-LSF algorithm has the highest score when the feature dimension is 570, with a score of 0.0924. In this case, better results can be obtained. In general, compared with other algorithms, SI-LSF has the best prediction effect and the best performance, and its prediction results are more effective. In practical applications, the SI-LSF algorithm has higher applicability with greater practicality. However, the selection of label-specific features does not take into account the shared information among variables, so it needs to be optimized for this aspect in the follow-up. At the same time, multi-objective regression algorithms have high applicability to practical application scenarios, so future research will consider applying the SI-LSF algorithm to practical problems in more fields, such as big data analysis, artificial intelligence, etc.

ACKNOWLEDGMENT

The research is supported by The first batch of construction projects of Applied Technology Collaborative Innovation Center of Fujian Higher Vocational Colleges (Min Jiao Ke [2016] No. 71 document);" Fujian Special Agricultural Products E-commerce Logistics Application Technology Collaborative Innovation Center" The science and technology plan projects of Fujian Provincial Department of science and technology (2022J01132808) "Construction of carbon footprint traceability and carbon label evaluation system of fresh agricultural products supply chain based on the blockchain technology". The science and technology plan projects of Fujian Provincial Department of science and technology (2023J011153) "Study on spatial and temporal

differentiation and development path of digital economy industrial structure in Fujian Province”.

REFERENCES

- [1] Chaudhuri KD and Alkan B. (2022). ‘A hybrid extreme learning machine model with Harris Hawks optimisation algorithm: an optimised model for product demand forecasting applications’, *Applied Intelligence*, Vol. 52, No. 10, pp. 11489-11505.
- [2] Dang Q L, Xu W, Yuan Y F. (2022). ‘A dynamic resource allocation strategy with reinforcement learning for multimodal multi-objective optimization’, *Machine Intelligence Research*, Vol. 19, No. 2, pp. 138-152.
- [3] Das T, Goerlandt F and Tabri K. (2022). ‘An optimized metamodel for predicting damage and oil outflow in tanker collision accidents’, *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, Vol. 236, No. 2, pp. 412-426.
- [4] Dong C. (2019). ‘Multiobjective optimization for unidirectional glass and carbon fiber-reinforced hybrid epoxy composites under flexural loading’, *Composites: mechanics, computations, applications*, Vol. 10, No. 1, pp. 39-68.
- [5] Grover A S and Syed A A. (2019). ‘Supply chain management & E-commerce: A review’, *India Quarterly*, Vol. 22, No. 4, pp. 5148-5153.
- [6] Han F, Zheng M and Ling Q. (2022). ‘An improved multiobjective particle swarm optimization algorithm based on tripartite competition mechanism’, *Applied Intelligence*, Vol. 52, No. 5, pp. 5784-5816.
- [7] Irodov V F, Barsuk R V and Chornomoret G Y et al. (2021). ‘Experimental simulation and multiobjective optimization of the work of a pellet burner for a tubular gas heater’, *Journal of Engineering Physics and Thermophysics*, Vol. 94, No. 1, pp. 219-225.
- [8] Jaipuria S and Mahapatra S S. (2021). ‘A hybrid forecasting technique to deal with heteroskedastic demand in a supply chain’, *Operations and Supply Chain Management an International Journal*, Vol. 14, No. 2, pp. 123-132.
- [9] Li C, Afak Yücel and Zhu K. (2017). ‘Inventory management in a closed-loop supply chain with advance demand information – ScienceDirect’, *Operations Research Letters*, Vol. 45, No. 2, pp. 175-180.
- [10] Li M and Zhang X. (2021). ‘Information acquisition and its incentives in an E-commerce supply chain under the offline showroom model’, *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 16, No. 5, pp. 1791-1804.
- [11] Masatoshi, Hatano and Toshifumi. (2020). ‘3-D shape recognitions of target objects for stacked rubble withdrawal works performed by rescue robots’, *Artificial Life and Robotics*, Vol. 25, No. 1, pp. 94-99.
- [12] Mirza T, Hassan M M and Hussain M W. (2020). ‘Indian Journal of Science and Technology Prediction of COVID-19 trend in India using time series forecasting’, *Indian Journal of Science and Technology*, Vol. 13, No. 32, pp. 3248 -3274.
- [13] Nabati M, Ghorashi SA and Shahbazian R. (2022). ‘JGPR: a computationally efficient multi-target Gaussian process regression algorithm’, *Machine Learning*, Vol. 111, No. 6, pp. 1987-2010.
- [14] Osojnik A, Panov P and Dzeroski S. (2017). ‘Multi-label classification via multi-target regression on data streams’, *Machine Learning*, Vol. 106, No. 6, pp. 745-770.
- [15] Pereira J, Francisco M B and Ribeiro RF et al. (2022). ‘Deep multiobjective design optimization of CFRP isogrid tubes using Lichtenberg algorithm’, *Soft Computing*, Vol. 26, No. 15, pp. 7195-7209.
- [16] Petkovi M, Kocev D and Deroski S. (2020). ‘Feature ranking for multi-target regression’, *Machine Learning*, Vol. 109, No. 6, pp. 1179-1204.
- [17] Puttagunta S, Pol C and Ferri M et al. (2020). ‘Diagnostic accuracy of single-phase computed tomography texture analysis for prediction of LI-RADS v2018 category’, *Journal of Computer Assisted Tomography*, Vol. 44, No. 2, pp. 188-192.
- [18] Ren J, Wei G and Bai J et al. (2020). ‘Target trajectory estimation by unambiguous phase differences from a single fixed passive sensor’, *Electronics Letters*, Vol. 56, No. 23, pp. 1270-1273.
- [19] Shen B and Chan H L. (2017). ‘Forecast information sharing for managing supply chains in the big data era: recent development and future research’, *Asia-Pacific Journal of Operational Research*, Vol. 34, No. 1, pp. 136-144.
- [20] Syed F H and Tahir M A. (2018). ‘Safe semi supervised multi-target regression (MTR-SAFER) for new targets learning’, *Multimedia Tools and Applications*, Vol. 77, No. 22, pp. 29971-29987.
- [21] Ullah W, Khan F and Umair M. (2021). ‘Multi-objective optimization of high torque density segmented PM consequent pole flux switching machine with flux bridge’, *China Electrotechnical Society Transactions on Electrical Machines and Systems*, Vol. 5, No. 1, pp. 30-40.
- [22] Wan Y M. (2021). ‘Amos-based risk forecast of manufacturing supply chain’, *International Journal of Simulation Modelling*, Vol. 20, No. 1, pp. 181-191.
- [23] Wang H and Jin Y. (2020). ‘A random forest-assisted evolutionary algorithm for data-driven constrained multiobjective combinatorial optimization of trauma systems’, *IEEE transactions on cybernetics*, Vol. 50, No. 2, pp. 536-549.
- [24] Xi X, Sheng VS and Sun B et al. (2018). ‘An empirical comparison on multi-target regression learning’, *Computers, Materials and Continua*, Vol. 56, No. 2, pp. 185-198.
- [25] Yang D, Xiao T and Choi T M et al. (2018). ‘Optimal reservation pricing strategy for a fashion supply chain with forecast update and asymmetric cost information’, *International Journal of Production Research*, Vol. 56, No. 5-6, pp. 1960-1981.
- [26] Wang X, Feng Z, Ying G. (2023). Emergency Parcel Dispatching and Structure Optimization of E-Commerce Logistics Network Based on CCNSGA-II. *Manufacturing and Service Operations Management*, 4(3): 50-56.
- [27] Moghadam S S, Aghsami A, Rabbani M. (2021). A hybrid NSGA-II algorithm for the closed-loop supply chain network design in e-commerce. *RAIRO-Operations Research*, 55(3): 1643-1674.

Construction and Application of Automatic Scoring Index System for College English Multimedia Teaching Based on Neural Network

Hui Dong*, Ping Wei

School of Foreign Languages, Tangshan Normal University; Tangshan Hebei, 063000, China

Abstract—With the continuous development of interactive multimedia, multimedia is increasingly integrated into college English teaching, providing advanced teaching equipment and resources. While enriching the teaching environment, it also brings new challenges to teaching ideas and strategies. Although the proportion of independent and selective learning of college students has increased, classroom teaching still constitutes the most essential unit of educational activities. Classroom evaluation is an important means and institutionalized element to improve the quality of university teaching. This paper analyzes the elements of multimedia classroom teaching and constructs an evaluation index system for English multimedia teaching. The improved model is used to achieve automatic teaching grading, acquire knowledge through environmental learning and improve its own performance, and evaluate the mathematical model of the English multimedia teaching evaluation system established by neural network theory accurately and effectively. In this paper, the results of automatic scoring of multimedia English teaching in colleges and universities are compared. Simulation software is used to verify the established neural network evaluation system. The simulation results show that the model is more suitable for the test data of English classroom teaching than the traditional methods, and the prediction effect is better. All 15 English teachers had a predicted error rate of less than 2%, and all 10 English teachers had a predicted error rate of less than 1%.

Keywords—Cognition of multimedia teaching in universities; scoring index; neural network; teaching system

I. INTRODUCTION

With the information in hypermedia structure to achieve human-computer interaction, multimedia college English classroom has become the driving force of classroom teaching mode reform, which has brought new challenges and development opportunities to teaching. In order to realize the hardware environment of multimedia classrooms, colleges and universities are committed to improving the level of multimedia teaching. How to scientifically apply multimedia technology to complete education is studied. The application has greatly changed the classroom teaching mode, such as promoting students' leading position and changing teaching strategies and methods. Modern educational teaching evaluation model guides and evaluates multimedia classroom teaching under the new situation [1]. The characteristics of multimedia teaching and the thought of modern educational technology are fully reflected in multimedia classroom teaching. Multimedia classroom teaching effectively promotes and determines the level of talent training and affects the

quality. Multimedia instruction needs new evaluation indicators to guide and promote education. The establishment of a multimedia classroom teaching index system is the guidance [2] Any evaluation system to develop scoring standards is the basis of the construction of the quality evaluation system of the discipline. By constructing the teaching quality evaluation system to form a closed-loop feedback mechanism, the regular operation and perfect development of each fundamental element of the system can be guaranteed to achieve the expected goal. Therefore, the evaluation system is the comprehensive embodiment of all levels and elements of the system, and the evaluation system is the system and operation mechanism. Consequently, it is of great significance to improve the evaluation of college classroom teaching, expand the management theory of college teachers and improve the quality of classroom teaching.

Neural network theory is the information science of human brain learning. A network is a multi-layer feedforward network in many types of neural networks. The neural network is widely used in decision analysis of nonlinear, complex and comprehensive problems [3]. Neural networks can be used as a qualitative and quantitative combination of practical tools to comprehensively evaluate the object system beyond the sample mode. The theory of artificial neural networks is a frontier field that learns from the human brain. The correct can be used as a qualitative and quantitative effective tool to make a comprehensive evaluation of the object system outside the sample mode. In the face of the number of thousands of learners, it is necessary to establish a clear standard assessment. Relying on machine learning automatic detection can accomplish this work while reducing teachers' workload. The algorithm is used in comprehensive universities. Establishing an evaluation system that adapts to the new teaching mode and conforms to the latest teaching concept has become an important topic.

Automated Essay Scoring (AES) was the first to start in the field of automated essay scoring. The Project Essay Grade (PEG) system can extract statistical information about the essay, such as the length of the text, the number of sentences, and the proportion of adjective prepositions. Then these statistical features are taken as multivariate independent variables, and the previously evaluated composition scores are taken as dependent variables. Logistic regression algorithm is used to find the representation function between these statistical features and the scores of English multimedia

teaching, and automatic scoring is achieved by finding out a multivariate function that can describe the relationship between these statistical information and the scores of English multimedia teaching. PEG system can achieve a high accuracy in composition grading, but because it cannot deeply understand English multimedia teaching, its grading results are difficult to be accepted later.

We can use natural language processing technology to develop areas of automatic grading of subjective questions to help teachers teach and students learn. Moreover, online teaching has become an important mode of current teaching. If AI-related technologies can be applied to English multimedia teaching, the model can be trained by using the students' homework that has been corrected online as training samples. Then, with the increase of training samples, the performance of the automatic scoring system will continue to improve, and the system will be closer to the level of teachers' manual scoring.

The rest of this article is organized as follows. Section II discusses the related work. Section III constructs an automatic scoring index system for multimedia English teaching. Section IV analyzes the application results of automatic scoring in college English multimedia teaching. Section V summarizes the full text.

II. RELATED WORK

The importance of English teaching evaluation has gradually attracted the attention of domestic college teachers and educational management departments, and the theory and practice of teaching evaluation have significantly been developed [4]. At present, domestic universities actively learn from the research results of computer networks. It takes multimedia technology and combines modern teaching models such as flipped classrooms, Massive Open Online Courses (MOOC), micro class and mobile device-assisted learning with traditional college English teaching. English through innovative teaching mode, and realize the transformation trend of professionalization, mobile, socialization and data. The primary methods of evaluation index systems include the brainstorming method, target element decomposition method, structure decomposition method and target classification method, and there have been rich theoretical research results. Based on the Analytic Hierarchy Process (AHP), Fu J et al. constructed a three-in-one inquiry classroom teaching evaluation system of teachers' teaching, students' learning and supervision evaluation [5]. Yang Y et al. pointed out that the classroom teaching evaluation system should include vector, positioning, concept, condition, operation and output subsystems [6]. Qilin S et al. established the monitoring and evaluation system, of course, teaching quality, including four processes of course application evaluation, Q evaluation system, early feedback of course teaching and stage evaluation, which involved course content, student development, teacher-student interaction, classroom organization, teaching effect, reading and homework [7]. Pustokhina I V et al. pointed out that the main focus of the change in the evaluation method was: the purpose and function of the evaluation should be to improve the happy atmosphere of the learning classroom and cultivate the self-confidence of students in physical learning [8]. Wang X et al. pointed out that the evaluation of teachers

should also involve students. Secondly, the evaluation should also pay attention to the combined evaluation of final results and usual performance, rather than one by one, and the evaluation process should also absorb students' participation [9]. Song R et al. pointed out online teaching, teaching quality, course content and other high-frequency words. It has been a hot issue in the field of online teaching evaluation recently. Evaluators often collect data on students' learning behaviours, preferences, motivations and other aspects through online teaching management systems, which provide important conditions and resource guarantees for learning analysis [10]. Kumar A et al. take the comprehensive evaluation problem of students' learning and examination scores as the evaluation object and use neural networks to establish a comprehensive evaluation model. Their research has achieved certain results and opened up a new path for teaching evaluation [11]. Based on the above considerations, the dynamic is the change in classroom teaching. The current classroom teaching evaluation still focuses on the static classroom teaching effect. Teaching quality includes teaching objectives, teaching attitude, teaching content, teaching process, teaching effect and so on, which cannot reflect the process and dynamic characteristics of classroom teaching. Media classroom teaching is a prerequisite for the creation of situations, the display of personality, the participation of learners and the development of learners' cognition. However, the current classroom teaching evaluation index is constantly improving and perfecting. The research on the assessment of English teaching needs to catch up. Educators need to construct a multivariate assessment model integrating the summative evaluation, diagnostic assessment and formative assessment, especially emphasizing the role of formative assessment. With the advantage of information technology, students' learning process is dynamically monitored, and the teaching quality is evaluated from the influencing factors.

III. AUTOMATIC SCORING INDEX SYSTEM FOR MULTIMEDIA ENGLISH TEACHING

A. English Teaching Effect Rating System

Based on the effect of systematic teaching from a multidimensional perspective, this paper constructs an English teaching effect scoring system and comprehensively considers it through measurable evaluation indicators. The evaluation system of English multimedia teaching should fully consider guiding the quality [12]. The index system discovers that the teacher puts forward the improvement direction by analyzing the existing problems in teaching. English multimedia teaching quality is decomposed into evaluation units from three structural elements: teaching management department, teacher performance and student response. Starting from the teaching management department, the standardization of teaching content is mainly considered. From the teachers' perspective, the evaluation especially involves teaching art and teaching attitude. From the perspective of students and learners, it primarily aims at students' reaction in the teaching process and evaluates the degree of their integration into classroom teaching. According to the expert survey method and feedback teaching theory, a number of measurable system sub-indicators are designed in detail, and the system of English classroom teaching effect is constructed, as shown in Fig. 1.

Classroom teaching is a process of dynamic change, reflected in each dynamic link, and each dynamic link's effectiveness is reflected to a certain extent [13]. It generally divides the per-preparation of teaching content, teaching art, teaching attitude, and student feedback. Teaching content is the key part of the design. Teachers clear the goal, and teaching is conducive to the direction. In the multimedia environment, the richness of resources is reflected in text materials, pictures, objects, audio, video and other aspects. Teachers should consider the support of various teaching resources when designing teaching. Under the support of teaching resources, teachers will create a particular situation according to the specific teaching content, such as the life as mentioned earlier situation, problem situation, simulated real-world situation, virtual situation, and other teaching resources that should be designed in connection with life. The observation point analysis of the multimedia classroom teaching process should start from the essential link of classroom teaching and determine the evaluation standard according to the goal that the teaching link should achieve.

B. Construction of Neural Network Classroom Teaching Evaluation Model

Neurons connect the different layers of the neural network, and neural networks can obtain a stable network system by repeated approximation learning [14]. The Back Propagation (BP) algorithm program updates the simple AHP classroom teaching evaluation index weight determination method. Model hierarchy is used to get the statistical sample data of the effect of classroom teaching evaluation, teacher evaluation sample validation set method divided into two parts of the training set and test set, enter a set of classroom teaching evaluation, the weighted average of the measure of a teacher's classroom teaching effect comprehensive score, as the training goals scored. The model calculates the neural network test score obtained by the investigated teachers, compares it with the target score of the sample data, and calculates the simulation

prediction error of the neural network statistical data. The estimated weights of the model are repeatedly debugged until a stable neural network is obtained. Based on the backpropagation method, a term is added to each weight and min value change proportional to the previous weight and min value change.[15]. The algorithm is as follows:

$$available\Delta X_i = \frac{-[J^T(x_i)J(x_i)+\mu I]^{-1}}{J^T(x_i)V(x_i)} \quad (1)$$

J is the Jacobian matrix, I is the identity matrix, the error vector, and V is the parameter vector. All inputs are submitted to the network, and the corresponding network outputs and errors are calculated using the following equation.

$$e_q = t_q - a_q^m a^0 = p \quad (2)$$

$$a^m = f^m(w^m a^{m-1} + b^m), m = 1, 2, \dots, M \quad (3)$$

Where 'p' is the input vector, 'w' is the weight vector of the MTH layer, 'f' is the vector of the transmission function of the MTH layer, 'a' is the output vector of the warp element generated in the MTH layer, and b is the bias value vector. For batch processing, the mean square error is the sum of the squared errors of all the targets in the training set [16].

$$the f(x) = \sum_{q=1}^Q \sum_{j=1}^s (e_j)^2 \quad (4)$$

The key in the algorithm is the calculation of the matrix. In order to compute the matrix, you need to replace the squared error's derivative with the error's derivative. The error vector is:

$$V^t = [e_{1,1}, e_{2,1}, \dots, e_{S^m,1}][e_{1,2} \dots e_{S^m,2}][e_{1,Q} \dots e_{S^m,Q}] \quad (5)$$

Where S^m is the total number of hidden layer nodes in the m layer.

$$N = QS^m, n = S_1(R + 1) + S_2(R^1 + 1) + \dots + S_m(R^m + 1) \quad (6)$$

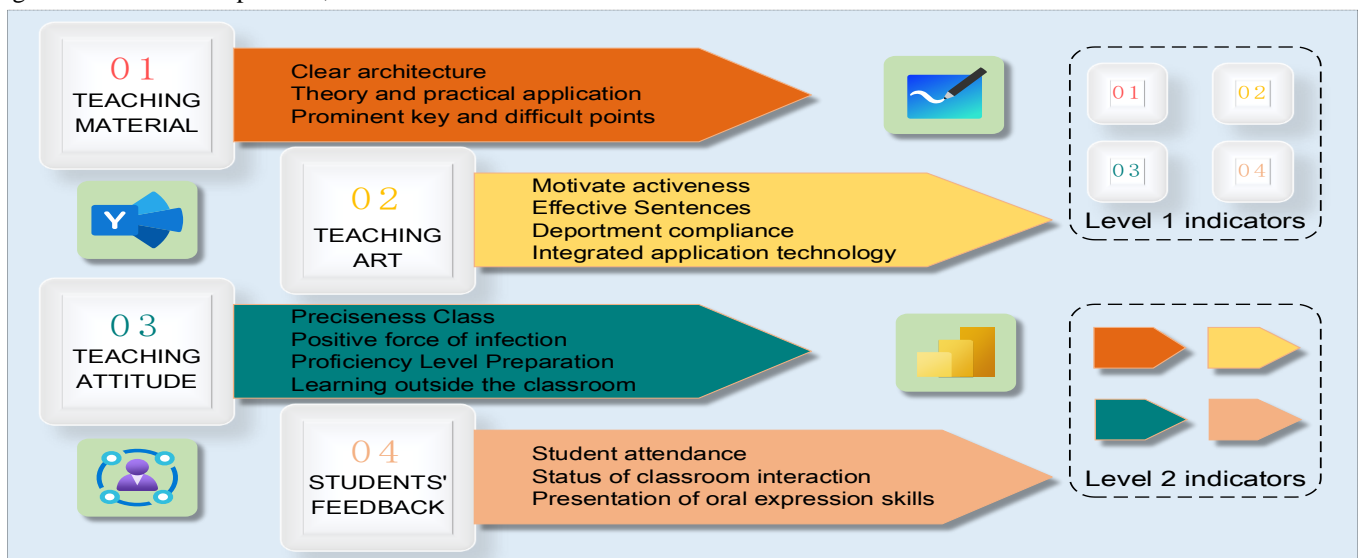


Fig. 1. Evaluation index system of English classroom teaching effect.

C. Improve the Learning Method of Neural Network

Neural networks acquire learning from the environment. The quality evaluation system of the teaching system starts from the main body and integrates the evaluation into any link. Only in this way can the evaluation system be formed to guide the whole system at a macro level. Each subsystem pays special attention to the combination of the teaching process and teaching results and the related factors affecting the teaching process and teaching results. Based on the above basic principles of the teaching system, this study proposes a roadmap of learning methods based on the quality evaluation system. In general, computational improvement is achieved gradually over time by adjusting its own parameters according to some predetermined measure [17]. The learning method provides information to supervise learning according to the environment, and the learning method is shown in Fig. 2.

In unsupervised learning, there is no external support in the supervision process. The learning system adjusts the training parameters uniformly according to the statistical law of the data provided by the external environment, and the structure is external fixed input. Supervised learning requires a model data set provided by the outside world. The input results correspond to the output results one by one, and this set of known data is set as the training sample set. The relearning system adjusts the system parameters according to the difference known. The external environment of relearning only gives evaluation information to the system output, but not the correct answer. Learning systems improve their performance by reinforcing actions. The quality evaluation system focuses on the formation of index content. In the thinking dimension of index content formation, the core subject of the system is sorted out, and the index content is clustered to form a particular general index type. The weight coefficient tests the evaluation system through teaching practice so as to create a perfect automatic scoring system in the repeated practice data.

D. Design of an Automatic Scoring System for Teaching Based on Improved Neural Network

Through continuous learning and training, artificial neural networks can discover their rules from complex data with unknown patterns, mainly can deal with arbitrary types of data. Therefore, this design will improve the theory of neural

network applied to the teaching evaluation system, solve the problem's evaluation index system, and overcome the traditional evaluation of the mathematical model and mathematical, analytical expression. Evaluate the accurate application of neural network theory to establish the mathematical model of the system [18]. If an error threshold exceeds the specified error range, adjust the connection weights among all layers and the threshold value of nodes. Therefore, the neural network is used in the comprehensive evaluation of the quality of teaching. The basic idea is to use each evaluation index of the neural network input vector with a value that is expert of the evaluation results of the neural network output vector. According to the comprehensive index system in the above section [19], the model is the required classroom model. The system design comprises various subsystems for guiding the ideology of network-based teaching within the respiratory field. These include the neural network evaluation subsystem for guiding teaching in the network, the multimedia resources neural network evaluation subsystem, the teaching conditions neural network evaluation subsystem, the teaching building neural network evaluation subsystem, the teaching management neural network study, the neural network evaluation subsystem for practice-based teaching in the building, and the English training neural network evaluation subsystem. The outputs of the eight subsystems constitute the integrated network's input system structure block diagram, as shown in Fig. 3.

Each subsystem model adopts a three-layer BP neural network, and neurons in each layer are only connected with neurons in neighbouring layers [19]. The conversion function of each node adopts the Sigmoid function.

$$S = \sqrt{0.43nm + 0.11m^2 + 2.55n + 0.77m + 0.35} \quad (7)$$

The index weights are obtained for each subsystem, and organizational education experts score the training samples according to the above index system. After training the model with a large number of samples, the subsystem is established, and the value is finally obtained. The model's structure of the system entails that the comprehensive output value of each subsystem is considered. The output is the teaching evaluation results divided into four grades: excellent, good, pass and fail. The value is shown in Fig. 4.

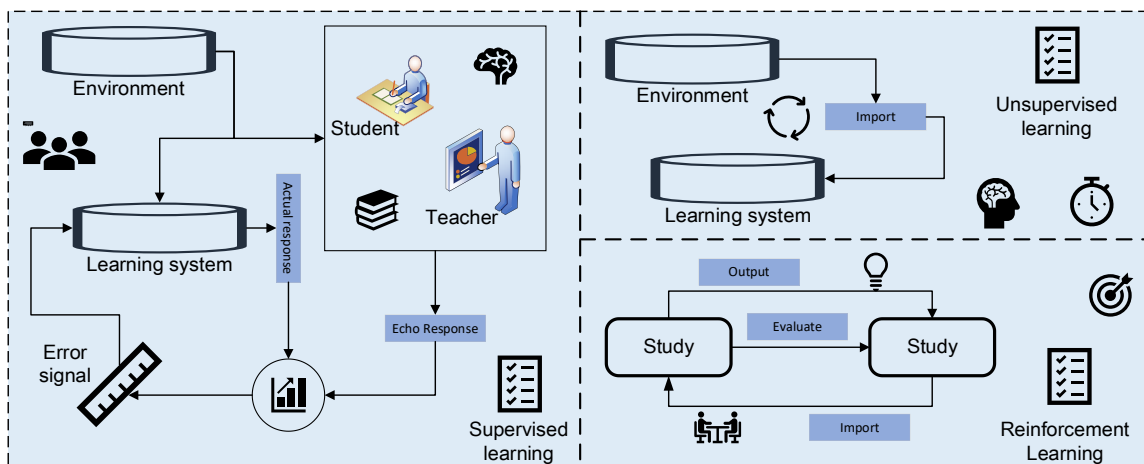


Fig. 2. Information content supervised learning.

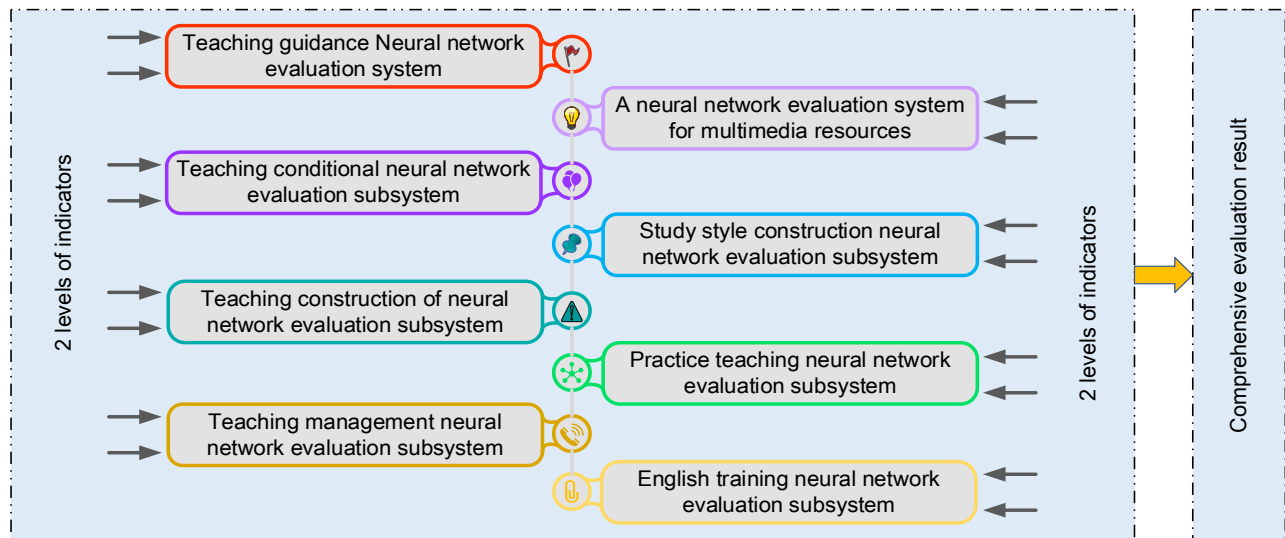


Fig. 3. Application flow chart of social security fund cloud audit platform.

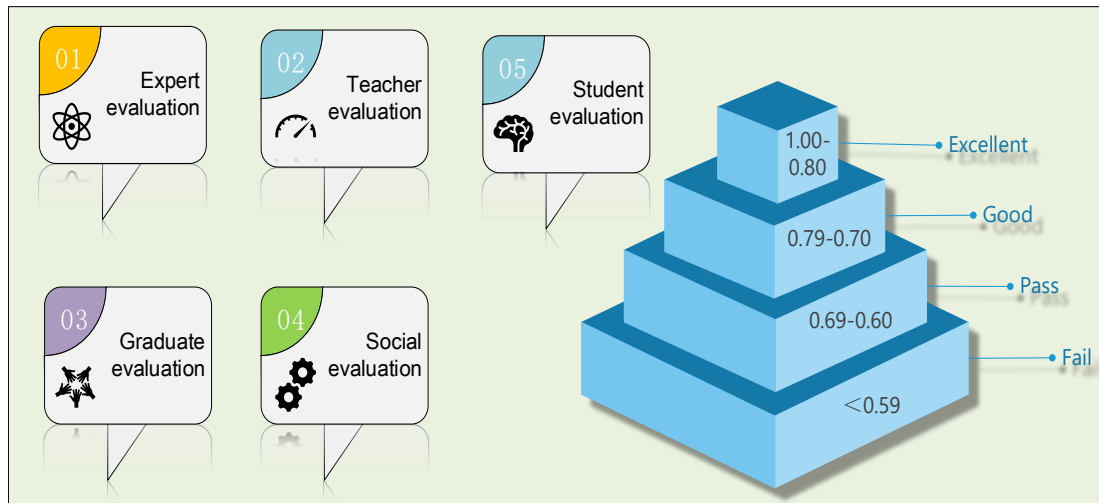


Fig. 4. Range of output values for each level.

The selection of neural network training samples is the key to this system and the evaluation results of the system. The system uses the evaluation of experts of the Ministry of Education, the evaluation of college teachers, the evaluation of students in school, the evaluation of graduates, social evaluation and other ways to obtain information using information cards and the Internet. The teaching evaluation score is between 1.00 and 0.80, which means the teaching activity is excellent. The teaching evaluation score is between 0.79 and 0.70, meaning the teaching activity is good. The teaching evaluation score is between 0.69 and 0.60, which means that the teaching activity is passed; A score of <0.59 in the teaching evaluation indicates that the teaching activity is failing.

IV. APPLICATION OF AUTOMATIC GRADING IN COLLEGE ENGLISH MULTIMEDIA TEACHING

A. Index Weight of the Automated Scoring System

In this study, the AHP is used to calculate the weight of indicators. AHP is an evaluation method that combines

quantitative analysis and qualitative analysis to simplify complex problems and simplify simple issues [20]. The process can effectively ensure the rationality and scientificity of index weight. The scoring matrices of 10 experts on the first-level indicators are summarized in a judgment matrix and summarized by the arithmetic average method, which is to take the average value of the values and calculates the corresponding summary value by the arithmetic average of the scoring values of each expert, as shown in Fig. 5.

The first-level index scoring judgment matrix was input, and the consistency test data was obtained, as shown in Fig. 6. In the consistency test, C.R. denotes the consistency ratio, C.I. denotes the consistency index, and R.I. indicates that the average consistency indicator is a fixed value. When $C.R. < 0.1$, the judgment matrix is consistent; If $C.R. > 0.1$, the judgment matrix does not meet the consistency requirements. If the data does not meet the requirements, delete the data, fill in the data matrix again, and calculate the weight after the consistency check.

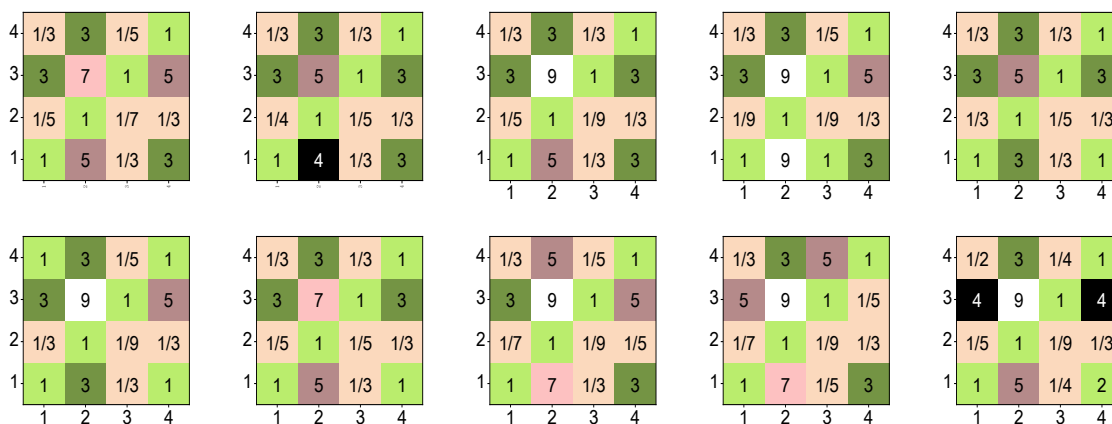


Fig. 5. Expert score arithmetic average summary numerical statistics chart.

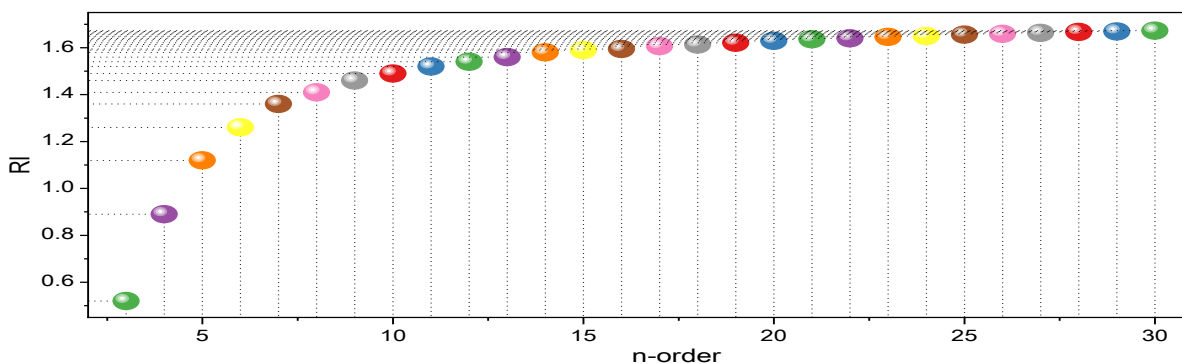


Fig. 6. Matrix consistency test data graph.

The C.R. value of the consistency test result is 0.044, indicating that the consistency test result meets the requirements and the calculated weight data. This process is effectively used to check the consistency of each expert's score. After passing the test, the arithmetic average method is used to carry out the weighted summary. The summarized values are re-entered into the table for the consistency test. After passing the consistency test, the weight value of each indicator is obtained. The formula is used for data entry, and the index weight is obtained. The single ranking of index weight

hierarchy means that the subordinate second-level indicators of a first-level indicator carry out the separate weight calculation, and the total ranking of hierarchy means that the second-level indicators of all first-level indicators carry out the comprehensive weight calculation and ranking. The results of the expert survey were analyzed through hierarchical analysis. The weights of the second-level indicators were obtained by calculating the data, and the consequences of the indicators were summarized as shown in Fig. 7.

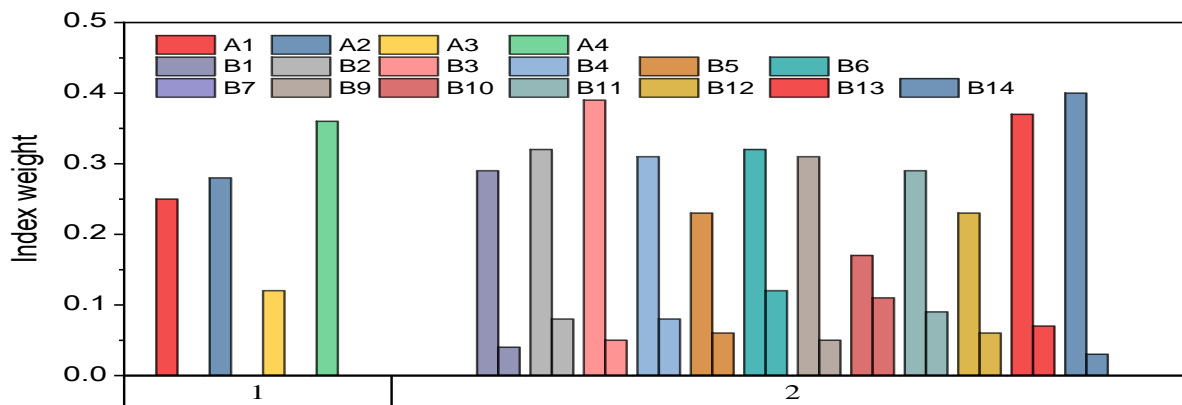


Fig. 7. Summary diagram of weights of system indicators.

This system designs eight subsystem output index weights. The weights are given 16 input values through model training, and then the teaching effect system evaluation results can be obtained. The evaluation is realized through the design of the evaluation database. From the database will be evaluated all teacher indicators into the system for database operation, evaluation of the main program neural network evaluation results display interface. The system monitors teaching, and the evaluation results can be used as a reference for teachers to conduct subsequent teaching activities. Multimedia teaching tools combine with teaching time to implement autonomous learning and offline face-to-face teaching organically. At the same time, learning should be combined with professional requirements, and the improvement degree of ability and quality should be matched with the learning objectives to improve students' satisfaction with course teaching.

B. Results of Neural Network Evaluation on the Quality of College English Multimedia Teaching

Matlab simulation's system design and application include efficient numerical calculation, matrix operation, signal processing, graph generation, and other functions. In addition, the software also provides a variety of practical toolboxes. Neural Networks Toolbox is one of them. Matlab neural network toolbox users can be very convenient for simulation. The system simulation was carried out in the toolbox, and the sample data were realized. It automatically mobilizes the initialization function and threshold based on the default parameters. Target vector T of the sample. According to the simulation steps, the system conducts simulation training and experimental analysis in the toolbox, and the results are shown in Fig. 8. According to the evaluation results of 10 groups of data by experts, it can be seen that the evaluation results of the neural network are consistent with those of experts.

The error of the test sample is almost the same as that of the test sample in the acceptable range. The statistical situation of items 4-8 in the questionnaire shows that 20% of the teaching simulation evaluation is excellent; 30% of the teaching simulation evaluation is good; 40% of the teaching simulation evaluation is passed; 10% of the teaching simulation evaluation

is failed. The error of the excellent evaluation test sample is 0.02, the good evaluation test sample is 0.01, and the error of passing the evaluation test sample is 0.01. The simulation error is within the controllable range. Using the automatic rating system to analyze the teaching quality, it is found that after the automatic rating system is enabled, the teaching performance only needs to input the corresponding index value and weight, and the evaluation subsystem can comprehensively evaluate the effect. Through teacher evaluation, students' learning effects can be objectively understood, and students' learning effects can reflect teachers' teaching effects. Therefore, the automatic evaluation model based on an improved neural network is a reasonable mechanical evaluation model. After the training evaluation is successful and the model is implemented, the repeated scoring process is separated from the experts by model learning. The application of the model not only saves capital investment but also ensures the scientificity of the model.

C. Evaluation Results of Neural Network Application in English Multimedia Teaching

The author obtains survey data and implements standardized preprocessing. This paper takes statistics teaching as the survey object, and the course is mainly offered for English majors. Utilizing the questionnaire survey, we obtain the classroom teaching data of some teachers in 10 universities, a total of 60 teachers. The sample data of all evaluation indicators were normalized, and the weight of scientific teaching content was the highest among all indicators. The weighted calculation obtained the classroom teaching quality score of all statistics teachers. The network structure takes 14 evaluation indicators as input neurons, the double hidden layer contains seven and three neurons, respectively, and the final output layer contains the unique prediction score of neurons. The error threshold of the model training was set as 10⁻⁵, and the times were set as 20000. The stable neural network was obtained by using the gradient descent algorithm. The neural network model predicts the classroom teaching scores of 15 teachers in the test set. The known classroom teaching scores are compared with the model, and the relative errors are calculated. The results are shown in Fig. 9.

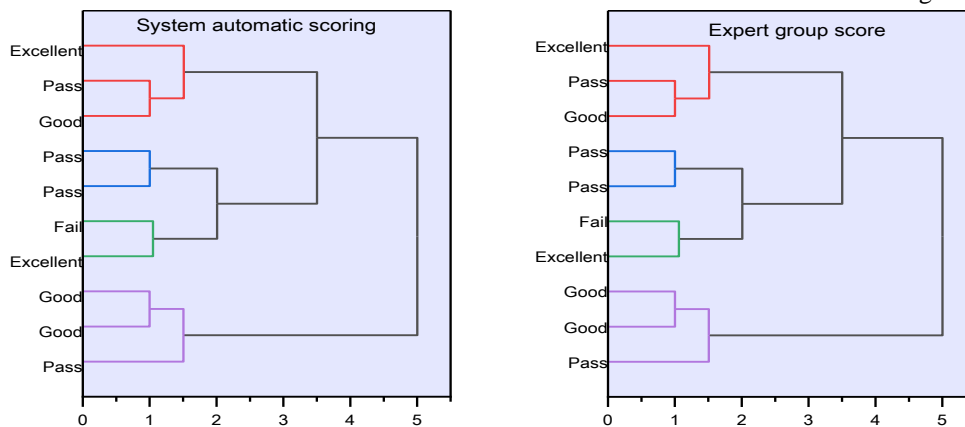


Fig. 8. Model simulation training results.

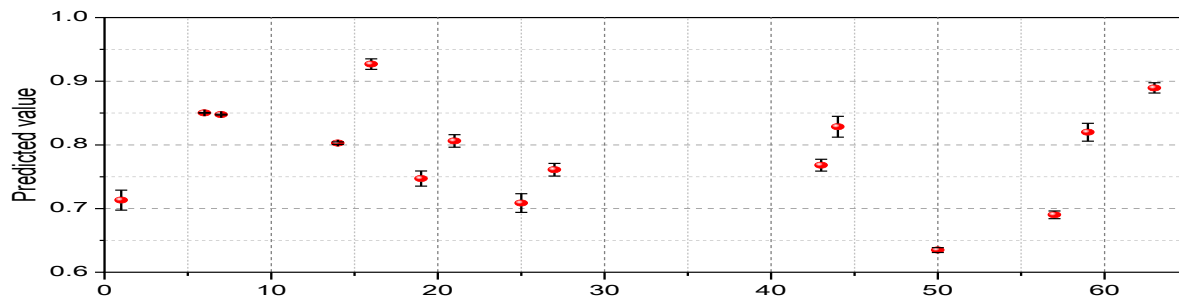


Fig. 9. Relative error between classroom teaching and neural network model prediction results.

The comparison of the above prediction results shows that the model is applied to the test data of English classroom teaching and obtains a better evaluation effect. The prediction error rate of the 15 teachers' classroom teaching quality score is controlled within 2%, and the error rate of 10 teachers is controlled within 1%. Therefore, the neural network teaching evaluation model can be put into the classroom teaching evaluation practice. Quantitative evaluation of teaching results and supervision of the process, and actively break through their own level in teaching. Evaluation promotes the formation of learning habits, according to learning needs to constantly adjust their learning state, always maintaining a serious and efficient learning state. Evaluations motivate teachers and students to strive for higher goals. A comprehensive review can effectively improve the teaching level to form good teacher supervision.

D. Discussion on Automatic Scoring of Multimedia English Teaching in Colleges and Universities

With the continuous increase of online teaching business and the rapid development of artificial intelligence technology, the current education model is developing in the direction of intelligence [21]. Online teaching will produce a lot of data, and the training of neural network model has a high demand for data [22]. Therefore, the application of neural network model to college English multimedia teaching is very innovative and practical value.

This paper studies the difficulties related to automatic online scoring of college English multimedia. The difficulty of automatic marking technology is mainly the correction of subjective questions, and the difficulty of subjective questions is the understanding of semantics. Subjective questions mainly rely on natural language processing, that is, text preprocessing, Chinese word segmentation, text vectorization, feature extraction and automatic scoring are carried out according to the processing order of natural language.

In this paper, by randomly combining the answers with the same score of the same question to form similar text pairs and positive sample pairs, the answers with large score differences or different questions are combined into negative sample pairs, so as to solve the problem of sample scarcity when using deep learning algorithms for automatic scoring.

At present, there are few researches on automatic scoring of Chinese subjective questions based on natural language processing, which has been the main research direction in the field of education [23]. In view of the importance of automatic scoring to promote the fairness of education and reduce the

redundancy of teachers' work, automatic scoring has begun to adopt the method based on neural network [24]. However, due to the different application fields of subjective questions, the emphasis of subjective questions in different professional fields is also different, especially the data of subjective questions does not have a relatively open and accurate data set. The current studies are all based on small-scale data from the school where the researcher is located, so it is necessary to construct a public data set of subjective questions.

V. CONCLUSION

The purpose is to provide feedback and teaching information to observe and promote the effect and progress of teaching, emphasizing improving the quality of teaching. College English teaching assessment encourages students to achieve their learning goals at all stages, cultivates their comprehensive English application ability, and helps teachers become conscious researchers of teaching theories and practitioners of teaching methods. This paper constructs the statistical index system of classroom teaching quality evaluation from teaching art, teaching attitude, students' reaction to classroom teaching and teaching content. Taking college English classroom teaching data as an example, this paper obtains the initial weight of evaluation. It calculates the target score of sample data by using expert scoring and an analytic hierarchy process. The neural network model trains the classroom teaching sample data to approximate the expert score, put into the test and inspection and obtains a good evaluation effect. The neural network model based on the BP algorithm makes use of the learning ability to maximize the inner connection between the teacher's teaching input information and the quality output, and it is separated from the expert scoring in the subsequent promotion and use. The sample size of the classroom teaching evaluation data set collected in this paper is small, and the advantages of the neural network model in extensive data analysis are more obvious if the model can be developed in a broader range of teaching activities. It is helpful to promote the transformation of teaching assessment form from results and figures to process and description, promote the combination of formative assessment and teaching, and improve the feasibility and scientificity of the assessment scheme.

REFERENCES

- [1] Y. Chen, "College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm," *J Sens*, vol. 2021, pp. 1–9, 2021.

- [2] M. Jin, "RETRACTED: Achievements analysis of mooc English course based on fuzzy statistics and neural network clustering," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 4, pp. 5559–5569, 2020.
- [3] G. Shi, X. Shen, F. Xiao and Y. He, "DANTD: A Deep Abnormal Network Traffic Detection Model for Security of Industrial Internet of Things Using High-order Features," *IEEE Internet of Things Journal*, 2023.
- [4] K. Li, X. Qian, and H. Meng, "Mispronunciation detection and diagnosis in l2 english speech using multidistribution deep neural networks," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 25, no. 1, pp. 193–207, 2016.
- [5] J. Fu, Y. Chiba, T. Nose, and A. Ito, "Automatic assessment of English proficiency for Japanese learners without reference sentences based on deep neural network acoustic models," *Speech Commun*, vol. 116, pp. 86–97, 2020.
- [6] Y. Yang and Y. Yue, "English speech sound improvement system based on deep learning from signal processing to semantic recognition," *Int J Speech Technol*, vol. 23, pp. 505–515, 2020.
- [7] S. Qilin, W. Xiaomei, F. Xiaoling, C. Yuanping, and W. Shaoyong, "Study on knee joint injury in college football training based on artificial neural network," *RISTI (Revista Iberica de Sistemas e Tecnologias de Informacao)*, no. E10, pp. 197–211, 2016.
- [8] I. V. Pustokhina et al., "Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems," *Ieee Access*, vol. 8, pp. 92907–92917, 2020.
- [9] X. Wang, D. Zhang, A. Asthana, S. Asthana, S. Khanna, and C. Verma, "Design of English hierarchical online test system based on machine learning," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 793–807, 2021.
- [10] R. Song, Z. Xiao, J. Lin, and M. Liu, "CIES: Cloud-based Intelligent Evaluation Service for video homework using CNN-LSTM network," *Journal of Cloud Computing*, vol. 9, pp. 1–9, 2020.
- [11] A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar, and M. Abdel-Basset, "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network," *IEEE access*, vol. 7, pp. 23319–23328, 2019.
- [12] W. Hui and L. Aiyuan, "A systematic approach for English education model based on the neural network algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3455–3466, 2021.
- [13] D. L. Minh, A. Sadeghi-Niaraki, H. D. Huy, K. Min, and H. Moon, "Deep learning approach for short-term stock trends prediction based on two-stream gated recurrent unit network," *Ieee Access*, vol. 6, pp. 55392–55404, 2018.
- [14] G. Shi, X. Shen, Y. He, and H. Ren, "Passive Wireless Detection for Ammonia Based on 2.4 GHz Square Carbon Nanotube-Loaded Chipless RFID-Inspired Tag," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, Art no. 9510812, 2023.
- [15] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput Educ*, vol. 147, p. 103778, 2020.
- [16] A. Korbach, R. Brünken, and B. Park, "Measurement of cognitive load in multimedia learning: a comparison of different objective measures," *Instr Sci*, vol. 45, pp. 515–536, 2017.
- [17] J. S. Mtebe, B. Mbwilo, and M. M. Kissaka, "Factors influencing teachers' use of multimedia enhanced content in secondary schools in Tanzania," *International Review of Research in Open and Distributed Learning*, vol. 17, no. 2, pp. 65–84, 2016.
- [18] J. Zhang and K. Yu, "Application of PPT playing system in the general city planning course under a multimedia teaching environment," *International Journal of Emerging Technologies in Learning (Online)*, vol. 11, no. 9, p. 36, 2016.
- [19] J. G. Smith and S. Suzuki, "Embedded blended learning within an Algebra classroom: a multimedia capture experiment," *J Comput Assist Learn*, vol. 31, no. 2, pp. 133–147, 2015.
- [20] L. McCoy, J. H. Lewis, and D. Dalton, "Gamification and multimedia for medical education: a landscape review," *Journal of Osteopathic Medicine*, vol. 116, no. 1, pp. 22–34, 2016.
- [21] Y. Wang, G. Gui, T. Ohtsuki and F. Adachi, "Multi-task learning for generalized automatic modulation classification under non-gaussian noise with varying SNR conditions", *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3587-3596, Jun. 2021.
- [22] R. Martinez-Maldonado, D. Gašević, V. Echeverria, G. Fernandez Nieto, Z. Swiecki and S. Buckingham Shum, "What do you mean by collaboration analytics? A conceptual model", *J. Learn. Analytics*, vol. 8, no. 1, pp. 126-153, 2021.
- [23] I-S. Comşa, G-M. Muntean and R. Trestian, "An innovative machine-learning-based scheduling solution for improving live UHD video streaming quality in highly dynamic network environments", *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 212-224, Mar. 2021.
- [24] Y. Lin, Y. Tu and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices", *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5703-5706, 2020.

Design of a Decentralized AI IoT System Based on Back Propagation Neural Network Model

Xiaomei Zhang

Marxist Academy, Henan Polytechnic Institute, Nanyang, 473000, China

Abstract—In the Internet of Things (IoT) era, when user needs are continually evolving, the coupling of AI and IoT technologies is unavoidable. Fog devices are introduced into the IoT system and given the function of hidden layer neurons of Back Propagation neural network, and Docker containers are combined to realize the mapping of devices and neurons in order to improve the quality of service of IoT devices. This study proposes the design of a decentralized AI IoT system based on Back Propagation neural network model. The testing data revealed that, at various data transfer intervals, the average transmission rate between the fog device and the sensing device was 8.265Mbps, and that the device's transmission rate could satisfy user demand. When the data transmission interval was 20s, the network data transmission rate was greater than 8.5Mbps and did not vary much when the number of data transmissions rose. The research demonstrates that the decentralized AI IoT system's network performance, which is based on a back propagation neural network model, can match user usage requirements and has good stability.

Keywords—BP neural networks; artificial intelligence; IoT systems; fog devices; Docker containers

I. INTRODUCTION

Artificial intelligence (AI) is a method for enhancing and extending human intelligence. It is a subfield of computer science that primarily focuses on the concepts of computer intelligence, human brain intelligent computers, etc. in order to advance computer applications [1]. The development of artificial intelligence (AI) can be aided by the mutual integration of mathematics and AI because AI is not confined to logical thinking but also requires the consideration of figurative and inspirational thinking [2]. The Massachusetts Institute of Technology was the first to suggest the idea of the Internet of Things (IoT), and the foundation of early IoT was radio frequency identification technology [3]. IoT applications now particularly encompass information interchange, sensing, and acquisition amongst objects. The seamless access to wireless networks has further broadened the field of IoT applications as a result of ongoing technological advancements [4]. The integration of IoT technology with computers, the Internet, and wireless communication technologies is the primary study area in the field of information technology today. IoT technology has become a popular field of technology because to its strong potential. IoT edge intelligence technologies and platform-based technologies for vertical applications are two examples of related technologies that have evolved as a result of the IoT technology boom [5]. The Internet of Things system is connected to sensors, actuators, and intelligent devices with

huge amounts of data. Usually, actuators only need data from local devices to respond, rather than all devices. All data is transmitted to the cloud, transmitting a large amount of inefficient data, resulting in a waste of network bandwidth. Sending sensor data to the cloud may introduce security vulnerabilities and privacy issues. The communication path from the terminal to the cloud is long and there are many nodes, making it susceptible to network attacks. The study recommends the Back Propagation Neural Networks (BPNN) model for decentralized AI IoT of Systems (IoTS) architecture in order to further improve the level of service offered by IoT devices. The study is broken down into four sections: a summary of current BPNN and IoT technologies; the design of decentralized AIoTS based on the BPNN model; an analysis of decentralized AIoTS applications based on the BPNN model; and a summary of the entire article.

II. RELATED WORKS

A multilayer feed forward network trained using the error back propagation algorithm; the BPNN is one of the most well-known neural network models. The results demonstrated that the particle swarm algorithm may increase the effectiveness of fault diagnosis. Xiao's scientific research team proposes a fault diagnosis system based on particle swarm optimization BP neural network for gearbox fault diagnosis. The particle swarm algorithm is used to optimize the weights and thresholds. The results show that the algorithm can improve the accuracy of fault diagnosis by up to 85% [6]. Yang in order to train the BPNN for the issue of information fusion state estimation of multi-sensor systems, the research team presented particle swarm and additional momentum approach. The simulation results prove that the method is effective [7]. In order to enhance the effectiveness of human shape prediction, Cheng used principal component analysis to reduce the dimensionality of pertinent variables. The experimental results show that the accuracy of this method is 12% higher than the K-means prediction model [8]. For the problem of enterprise asset valuation, Xie's group proposed a BPNN-based valuation model for technology enterprises. The model included financial and non-financial performance indicators related to intellectual property, and the results indicated that the inclusion of these indicators could enhance the model training effect [9]. Shi proposed a BPNN-based short-term load forecasting model for smart grids, in which various types of data are fed into the model and its output is represented as conforming to the forecast results. Experiments show that the method is able to clearly display the distribution of load demand at various time periods [10].

The IoT architecture can be divided into a sensing layer, a network layer, and an application layer. It uses communication technologies like local networks or the Internet to connect sensors, controllers, machines, and people in novel ways to create intelligent networks that connect people to things and things to people. Using a near real-time approach to data streams and big data-based pattern analysis of stakeholder needs, Luckner and his team propose an IoT architecture that is data-centric for urban services and applications, and test results show that the approach is effective at lowering latency in smart city IoT [11]. In order to ease further analysis of human health, Chandrakar's research team presented a smart device for healthcare system based on IoT architecture employing smart sensors for human tracking. The study's findings show that the technique can support IoT architecture for healthcare [12]. The research group of Shapsough suggested a general IoT architecture for context-aware learning. To enable on-the-fly scene learning, a variant of the architecture is constructed utilizing IoT edge devices, and testing findings demonstrate that the architecture is resource-efficient while limiting application protocols [13]. In order to meet the demands of the IoT ecosystem, the Sarrigiannis scientific group built a 5G platform using virtual network capabilities for lifecycle management of heterogeneous architectures in conjunction with multi-access edge computing. Experiments reveal that the method can allocate edge and core resources in real time to maximize the number of service users [14]. In order to connect travel paths, Cheng and his team proposed a new carrier-based sensor deployment algorithm that matches redundant sensors with uncovered areas. Experimental data shows that under different parameter settings, this algorithm can reduce the path length of rows by 20% -30% [15].

In conclusion, BPNN and IoTS structures have been the subject of several studies and designs by numerous research teams, but more work has to be done to increase their stability. The study recommends a decentralized AIIoTS design built upon the BPNN model in order to boost the data transmission rate.

III. DECENTRALISED AIIoTS DESIGN BASED ON BPNN MODEL

The design of the AIIoTS in this chapter makes use of BPNN. The architecture of the system based on the BPNN model is covered in the first half of this chapter, and the implementation of the decentralized AIIoTS capability is covered in the second section. To implement the mapping of BPNN and IoTS, Fog Devices (FD) and Docker Containers (DC) are introduced.

A. System Design based on the BPNN Model

BPNN is one of the widely used and more successful neural networks [16]. When the signal is propagated forward, the data flows from the input layer (IL) to the hidden layer (HL), and the error is back-propagated by comparing the actual value with the expected value in the output layer [17]. Fig. 1 depicts the BPNN's construction.

The IL, output layer and HL form the BPNN, and the forward propagation of the BPNN algorithm is calculated as shown in Eq. (1).

$$\begin{cases} net_{ij} = W_{(i-1)kj} * \sum_{k=1}^{N_{i-1}} O_{(i-1)k} \\ O_{ij} = \frac{1}{1 - \exp[-net_{ij} + \theta_{ij}]} = f_s(net_{ij}) \end{cases} \quad (1)$$

In Eq. (1), the total input, output and threshold of the j th neuron in layer i are net_{ij} , O_{ij} and θ_{ij} respectively, the number of neuron nodes in layer i is N_i , and the connection weight of the j th neuron in layer i to the k th neuron in layer $i+1$ is W_{ijk} . The error in the backoff algorithm is defined as shown in Eq. (2).

$$e_j = d_j - y_j \quad (2)$$

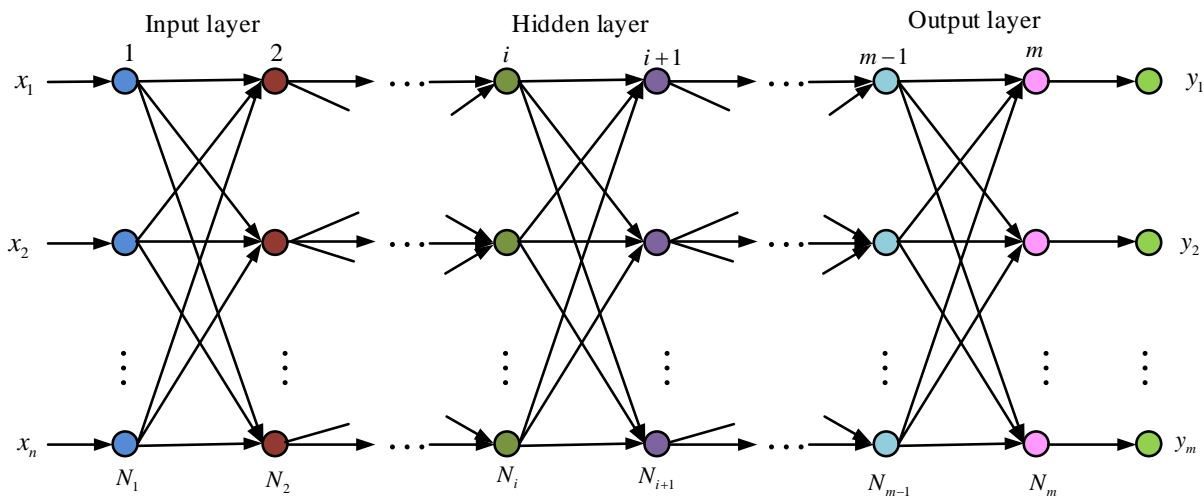


Fig. 1. BP neural network structure.

In Eq. (2), the back propagation error is e_j , the desired output is d_j , and the actual output of the neural network is y_j . The network objective function is shown in Eq. (3).

$$E = \frac{1}{2} \sum_j (d_j - y_j)^2 \quad (3)$$

In Eq. (3), the network objective function is E , and the weights are calculated as shown in Eq. (4) with the correction value along the direction of the gradient of the objective function falling.

$$\Delta W_{ijk} = -\eta \frac{\partial E}{\partial w_{ijk}} \quad (4)$$

In Eq. (4), the gradient descent value is ΔW_{ijk} and the learning efficiency is η , which takes values in the range of [0,1]. The recursive relationship between the gradient descent value and the neuron output is shown in Eq. (5).

$$\Delta W_{ijk} = -\eta \frac{\partial E}{\partial w_{ijk}} = -\eta \frac{\partial W}{\partial net_{(i+1)k}} * \frac{\partial net_{(i+1)k}}{\partial w_{ijk}} = \eta \delta_{ik} \frac{\partial net_{(i+1)k}}{\partial w_{ijk}} \quad (5)$$

In Eq. (5), the whole is denoted by δ_{ik} in terms of $\frac{\partial E}{\partial net_{(i+1)k}}$.

The common principles of BPNN design are the selection of the input quantity that will be able to meet the feature reflection requirements; the output quantity is the target that the system needs to reach; the training set samples can meet the generalization ability of the test metrics; the initial weight setting needs to meet the initial net input of the nodes as close to zero as possible; the number of HL neurons is calculated is shown in Eq. (6).

$$\begin{cases} m = \sqrt{n+l} + \alpha \\ m = \sqrt{nl} \end{cases} \quad (6)$$

In Eq. (6), m stands for the number of HL neurons, l for the output neurons, n for the input neurons, and α for the regulatory constant, which has a range of values from [1,10]. The study uses the BPNN structure as a model to construct the decentralised AIIoTS, and the whole system is divided into the cloud computer layer, and the decentralised AIIoTS architecture is shown in Fig. 2.

The edge of the IoT is the end device layer, containing IoT-aware devices and execution devices, mapped to the IL and output layers of the BPNN respectively, with one device corresponding to one neuron. The IoT-aware devices send the collected data information to the fog processing layer. The middle layer of the cloud hybrid architecture is the fog processing layer. The cloud computing layer, which makes up the top layer of the entire architecture, principally consists of cloud servers in charge of storing data, training neural networks, and distributing devices. The decentralised AIIoTS architecture works in both decentralised and centralised ways. Decentralized working involves the absence of the cloud

computing layer, the deployment of the BPNN's input devices as sensing devices, the HL deployment of the fog processing layer, and the output devices as execution devices. The cloud computing layer is required for centralised work and its function is to train the neural network, which is a component of the BPNN algorithm deployment. The study uses a three-layer BPNN as a model, with HL neuron inputs as shown in Eq. (7).

$$net_j = \sum_i W_{ij} O_i \quad (7)$$

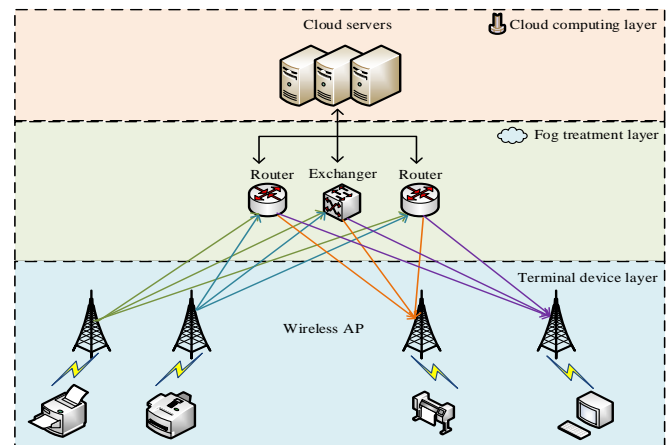


Fig. 2. Decentralized Artificial intelligence of things architecture.

In Eq. (7), the input of the j th neuron of HL is net_j , the weight of the i th neuron of IL and the j th neuron of HL is W_{ij} , and the output of the i th neuron of IL is O_i . The output of the HL neuron is shown in Eq. (8).

$$\begin{cases} O_j = g(net_j) \\ g(x) = \frac{1}{1 + e^{-(x+\theta)}} \end{cases} \quad (8)$$

In Eq. (8), the output of the j th neuron of HL is O_j and the threshold is θ . The IL and HL neuron weights change as shown in Eq. (9).

$$\Delta W_{ij} = \eta O_j (1 - O_j) \sum_k \delta_k W_{kj} O_i \quad (9)$$

In Eq. (9), the weight change of the i th neuron of IL and the j th neuron of HL is ΔW_{ij} , the output of the k th neuron of HL is W_{kj} , the residual of the k neuron of the output layer is δ_k , and the learning efficiency is η . The threshold change of the HL neuron is shown in Eq. (10).

$$\Delta \theta_j = \eta O_j (1 - O_j) \sum_k \delta_k W_{kj} \quad (10)$$

In Eq. (10), the threshold change for the j h neuron of HL

is $\Delta\theta_j$, and the output layer neuron input is calculated as shown in Eq. (11).

$$net_k = \sum_j W_{kj} O_j \quad (11)$$

In Eq. (11), the input of the k th neuron in the output layer is net_k , and the output of the neuron in the output layer is calculated as shown in Eq. (12).

$$O_k = g(net_k) \quad (12)$$

In Eq. (12), the output of the k th neuron in the output layer is O_k , and the weights between the output layer and HL neurons change as shown in Eq. (13).

$$\Delta W_{kj} = \eta(t_k - O_k) O_k (1 - O_k) O_j \quad (13)$$

In Eq. (13), the weight of the k th neuron in the output layer is changed to ΔW_{kj} with the j th neuron in HL, and the residuals of the neurons in the output layer are shown in Eq. (14).

$$\delta_k = (t_k - O_k) O_k (1 - O_k) \quad (14)$$

In Eq. (14), the residual of the k th neuron in the output layer is δ_k , the expected output of the k th neuron in HL is t_k , and the threshold of the neuron in the output layer changes as shown in Eq. (15).

$$\Delta\theta_k = \eta(t_k - O_k) O_k (1 - O_k) \quad (15)$$

In Eq. (15), the threshold change for the k th neuron of the output layer is $\Delta\theta_k$. The FD layer is given the function of an HL neuron, and forward and backward calculations are required during training.

A. Decentralised AIoTS Functional Implementation

Once the system based on the BPNN model is constructed, it needs to be implemented in conjunction with the appropriate tools and devices. The study uses the Python programming language and deploys DCs on FD and execution devices. The Python programming language is simple, easy to learn and implement, and is a free and open source software with source code that can be read and modified [18]. Python is a high-level language that is programmed without the need to consider low-level details when programming, and has the advantages of portability, interpretability and extensibility [19]. Docker's object is server-side and belongs to a Linux container technology, Docker can be installed on most Linux systems [20]. The first step in the study to install DC on FD and execution devices is to check the kernel version. 64-bit computers are required to install and run Docker, so the Linux system kernel version needs to be greater than 3.0, and the kernel needs to be upgraded if it is not up to standard [21]. The second step is to update the advanced packaging tool source, and the third step is to install the DC according to the command. The Docker image is the basis for the DC build,

and the study builds the image using a Dockerfile file. The mapping of the IoT nodes to the BPNN, the code written in the DC, gives the FD and the execution device the function of hiding neurons and output neurons, respectively. Fig. 3 displays the decentralized AIoTS capabilities based on the BPNN model.

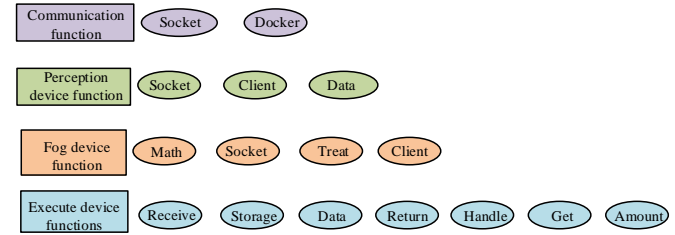


Fig. 3. Functions of decentralized Artificial intelligence of things system based on BP neural network model.

Communication between DCs needs to be implemented using sockets, also called sockets, which support the TCP/IP network communication protocol and are the endpoints for bidirectional communication between different hosts [22]. There are three types of sockets. Streaming sockets use the TCP protocol, which provides a reliable connection-oriented bi-directional data transfer service that guarantees error-free data transfer and is suitable for high-volume and data transfer demanding communication situations. Datagram sockets use the UDP protocol, which provides a connectionless service that does not guarantee reliable and sequential data transfer and requires programmatic processing to be used. Raw sockets use the underlying protocol and are suitable for network protocol analysis and verification [23]. Socket communication requires the availability of a server and a client, and the connection process consists of server listening, client request and connection confirmation. The role of the client is to connect to the server and send data to the server side.

The study uses the socket server module to simplify the web application, which contains a framework of service classes and request processing classes. The first step in the creation of the service is the creation of the service class, which uses the TCP protocol to enable asynchronous processing [24]. The second step is the creation of the request processing class, along with the overriding of the processing functions. The third step is the instantiation of the service object, for which the service address and request processing class are passed. The fourth step is the invocation of the service class object function and the server is kept running after the function is started.

The inverse device is used to acquire data, which is then sent to the fog processing layer, which contains the hardware software study of the sensing device [25]. This study only simulates the function of the device; it does not create a functional implementation. Instead, it uses the Client function, which is the Client function of the upper level, to send data. The function of FD is to process, store and send the data, and its DC is distributed in the server side and the client side, IL and Output layer client request and processing is completed through the server side, FD function implementation flow is shown in Fig. 4.

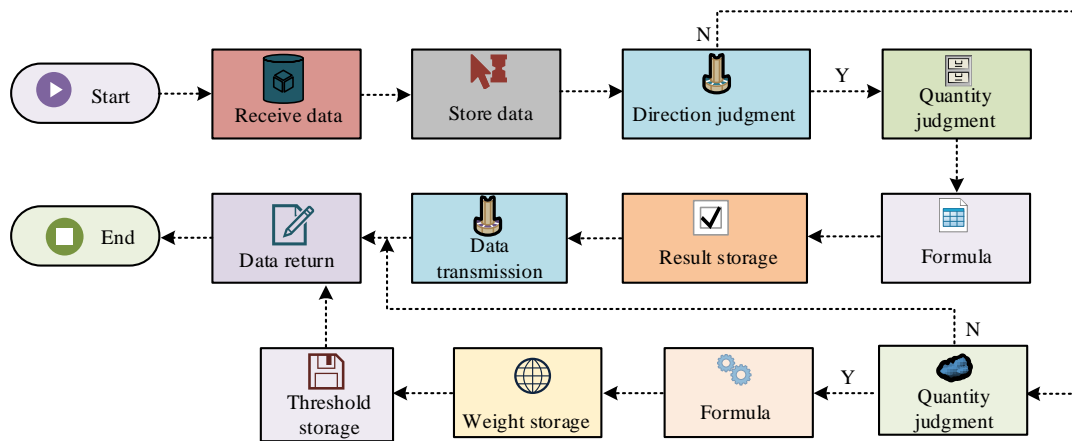


Fig. 4. Implementation process of fog equipment functions.

FUNCTIONS USED TO IMPLEMENT DEVICE FUNCTIONS

Function Name	Parameter	Return Value	Function
Client	Host, port, senddata	/	Connect to the server of the hidden layer and send data
Tansform	Filename, number	Data_list	Convert the relevant data in the file into a list
Error_term	D, output	Output_e	Calculate the residual of a single output layer neuron
Th_updata	Output_e, efficiency, th	Updatated_th	Update threshold
Save_data	Data, number	/	Store data in the appropriate location
Get_Data	Filename, number	Data	Obtain data with corresponding numbers from files containing storage numbers
Handle	Self	/	Please handle hidden layer clients Request, process received data
Freedforward	F_input, w, th	Output	Forward calculation of output layer
Weight_update	Output_e, efficiency, f_input, w	Updatated_weight	Update the weights of output layer neurons to corresponding hidden layer neurons
Save	Filename, data	/	Store data
Get_number	Data	Number	Obtain storage number
Amount	Data, number	Data_number	Obtain the number of stored data corresponding to the current storage number

The first step in the implementation of the FD function is the reception and storage of data, which comes from the sensing and execution devices. The format of the data storage is JSON, which is a lightweight data exchange format. The second is to process the data, firstly to determine the type of data, forward data from the sensing device and backward data from the executing device. Returning the data to the connected client is the third stage, and whether data processing is done or not has no bearing on the client obtaining the returned data. The functions used to implement the functions of the execution device are shown in Table I.

The main function of the execution device is also data processing, storage and sending, its DC is distributed in the client and server side, the source of processing data is FD, the format of data storage is still JSON format, the storage file is f_input.json, data processing before also to determine whether to meet the requirements, the number of data and the number of hidden neurons equal to the judgment criteria, after receiving data are returned a data to the client.

IV. ANALYSIS OF DECENTRALISED AIIoTS APPLICATIONS BASED ON THE BPNN MODEL

This chapter addresses the application and analysis of decentralised AIIoTS based on the BPNN model. The first section of this chapter is a simulation analysis of the performance of decentralised AIIoTS, and the second section of this chapter is an analysis of the practical application of decentralised AIIoTS.

A. Performance Simulation Analysis of Decentralized AIIoTS based on BPNN Model

Two laptops were used to simulate the FD and IoT sensing devices, with Laptop 1 simulating the FD and Laptop 2 simulating the sensing device. The operating system of both computers is Ubiquitous 20.10, and the network connection of laptop 2 is provided by the hotspot of laptop 1. The DCs are deployed at the corresponding locations of laptops 1 and 2, and the FD transmission rate is shown in Fig. 5.

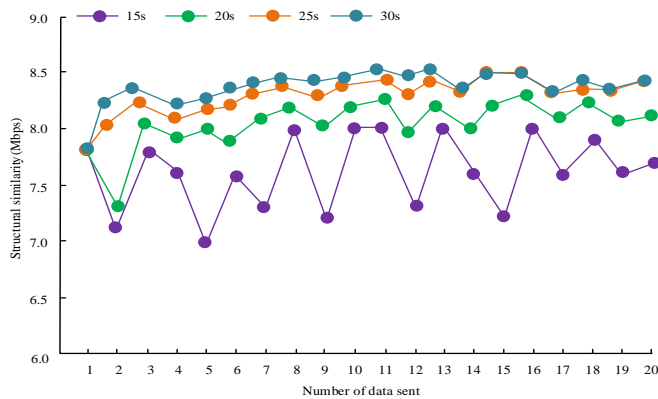


Fig. 5. Transmission rate of fog equipment.

In Fig. 5, laptop #2 sends data to laptop #2 at 15s, 20s, 25s and 30s intervals, and 20 times at different time intervals. It can be seen that the transmission rate between FD and sensing device is distributed in the interval [7.0,8.5] Mbps, with an average transmission rate of 8.265 Mbps, and the transmission rate increases when the data transmission interval increases. The results show that the transmission rate of the device under the decentralized AIIoTS based on the BPNN model proposed in the study can meet the user requirements. To further analyse the communication link bandwidth (marked as BP) under the decentralized AIIoTS based on the BPNN model, the experiments used the communication link bandwidth between the common sensor and the cloud server (marked as S), the Unified Storage Network architecture link communication bandwidth (marked as U), and the machine-to-machine architecture link communication bandwidth (marked as M) as comparisons, and the bandwidth comparison results are shown in Fig. 6.

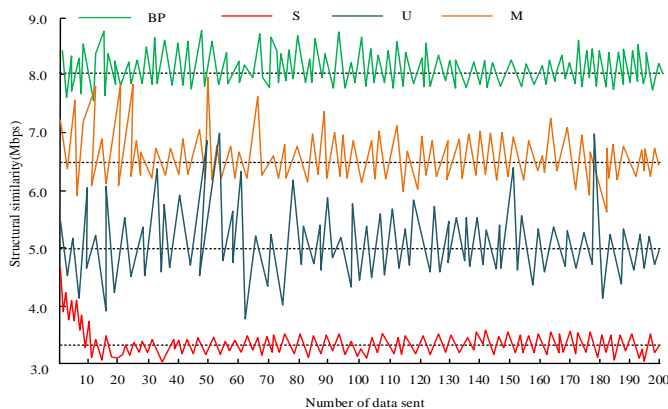


Fig. 6. Bandwidth comparison results.

In Fig. 6, the data is sent at an interval of 10s and a total of 200 times. It can be seen that the average communication link bandwidth between ordinary sensors and cloud servers is 3.2Mbps, the average link communication bandwidth of the Unified Storage Network architecture is 5.0Mbps, the average

link communication bandwidth of the machine-to-machine architecture is 6.43Mbps, and the average communication link bandwidth under the decentralized AIIoTS based on the BPNN model is 8.01Mbps. The outcomes demonstrated that the study's decentralized AIIoTS may offer greater communication bandwidth, faster data transmission rates, and higher stability. A comparison of the data transmission delay under decentralized AIIoTS and the delay of sensor data sent to the cloud is shown in Fig. 7.

In Fig. 7, a total of six sensing devices are set for data transmission. When the number of sensing devices is 1 and 2, the difference in latency between the data sent by the sensing devices to the FD and the cloud is small, and the difference in latency gradually increases, the latency of sending data to the cloud is 1.80s, the latency of sending data under decentralized AIIoTS is 0.83s, and the latency is reduced by 0.97 s, and the latency of sensor data sent to the cloud is 1.74s, with a latency reduction of 0.06s. The results show that the latency reduction of data sent under the decentralized AIIoTS based on the BPNN model proposed in the study is greater. To further validate the correct data delivery rate under the decentralized AIIoTS (labelled as BP), the study uses the Unified Storage Network architecture (labelled as U) and the machine-to-machine architecture (labelled as M) as comparisons, and the correct data delivery rates of the different architectures at different data delivery intervals are shown in Fig. 8.

In Fig. 8, three data transmission intervals are set to 20s, 15s and 10s, respectively. It can be seen that the correct data transmission rate for decentralized AIIoTS and with a data transmission interval of 20s is 1.00, and the correct data transmission rate cannot be maintained at 1.00 under the Unified Storage Network architecture and machine-to-machine architecture. The experimental results show that the data processing and transmission are better under the decentralized AIIoTS based on the BPNN model proposed in the study, and the data sending interval can be set to 20s if there is no special requirement.

B. Analysis of Decentralised AIIoTS Applications based on the BPNN Model

To verify the effectiveness of the practical application of decentralized AIIoTS based on the BPNN model, six users were experimentally recruited to analyse the network performance under the system, setting the data transmission interval to 20s and the user network data transmission rate as shown in Fig. 9.

As shown in Fig. 9, none of the users' network data transmission rates are less than 8.5Mbps, and even as the volume of data transmissions rises, these rates remain stable in the [8.4,8.7] Mbps range, which can accommodate users' typical usage requirements. When the data transmission interval is set to 10s, the user network data transmission error is shown in Fig. 10.

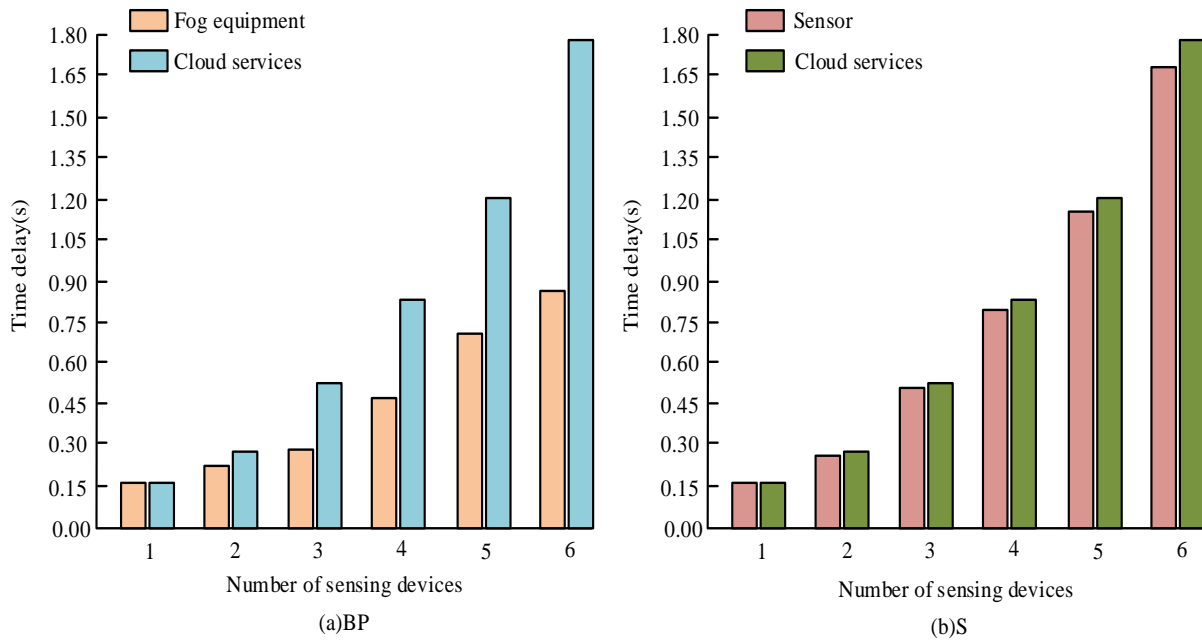


Fig. 7. Comparison of data transmission delay.

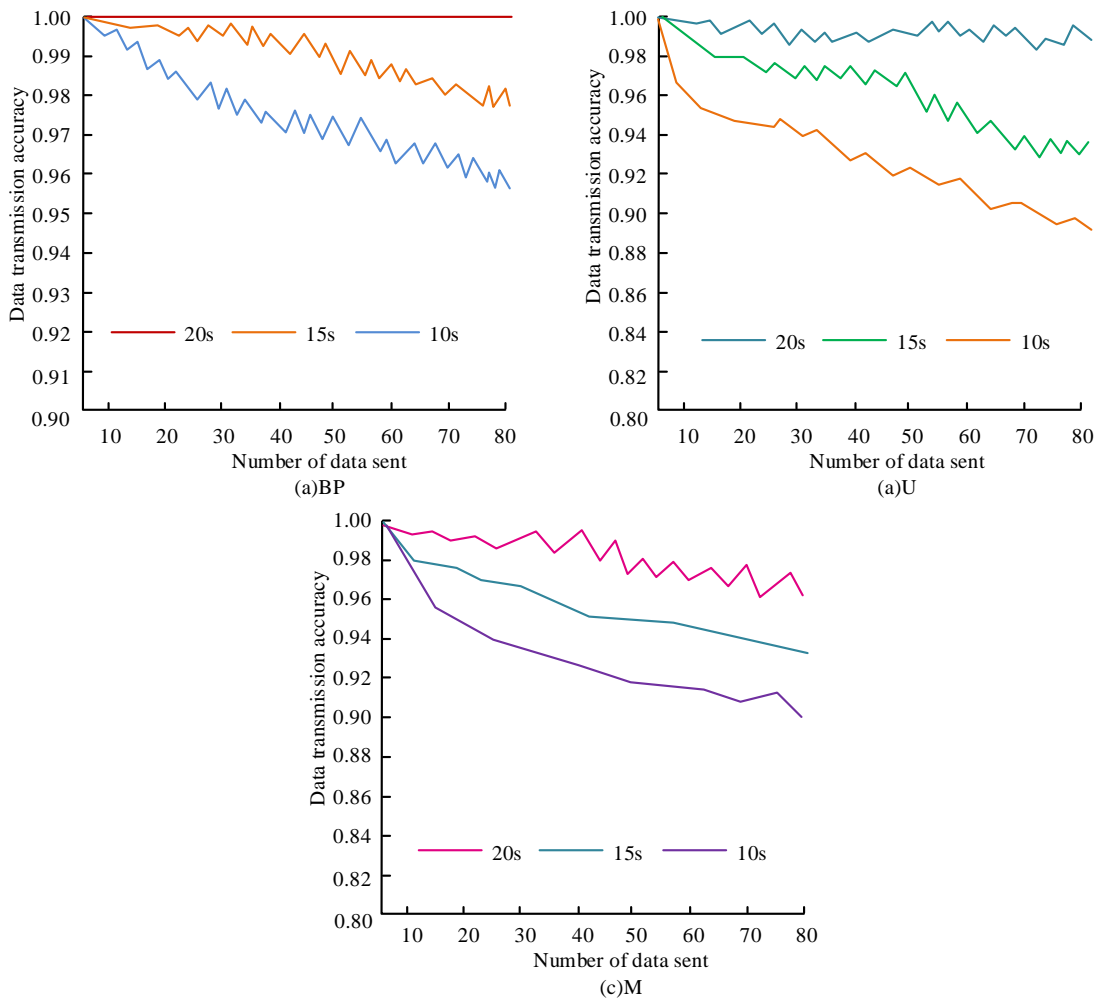


Fig. 8. Data transmission accuracy of different architectures at different data transmission time intervals.

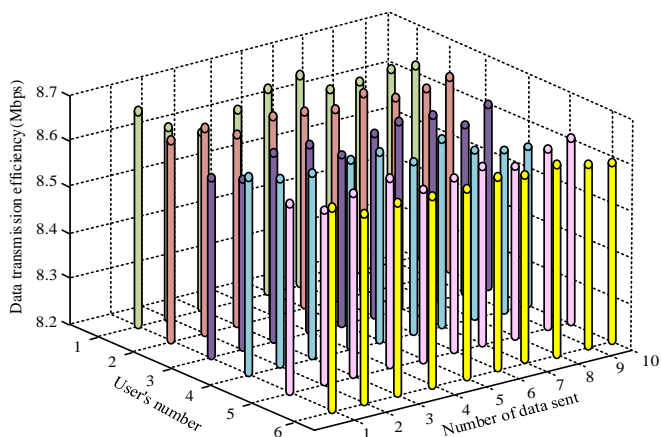


Fig. 9. User network data transmission rate.

Fig. 10 illustrates how the network data transmission error increases gradually as the number of users rises, although the error is still less than 0.07. Analysis of the experimental data shows that the correct rate of network transmission at this time is higher than 0.93. The accuracy of user data transmission can be guaranteed when there are no special requirements for user data transmission accuracy. To verify the security of the network transmission under the decentralised AIoTS based on

the BPNN model, the experiments were verified using an attack test, and the results of the attack test are shown in Fig. 11.

In Fig. 11, the network is able to identify 85% of threat attacks when the data transmission interval is 10s, 95% of threat attacks when the data transmission interval is 15s, and 100% of threat attacks when the data transmission interval is 20s. The experimental data shows that the security of network data transmission under the decentralised AIoTS based on the BPNN model is high and can meet users' needs.

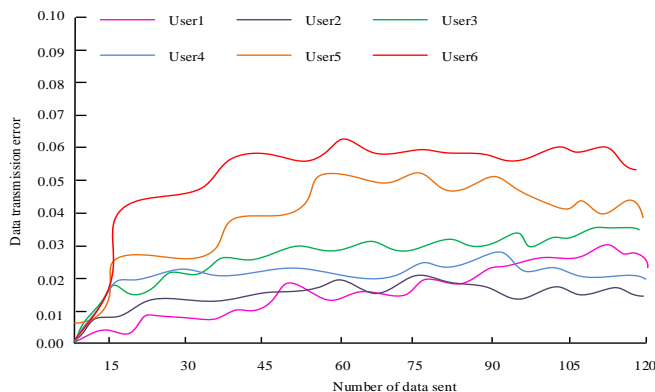


Fig. 10. Data transmission error.

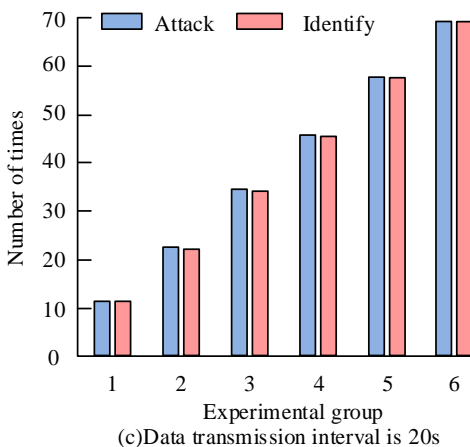
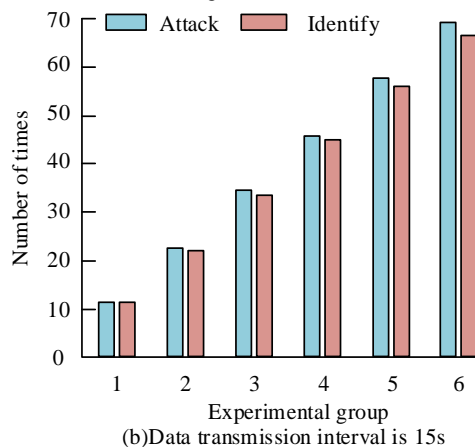
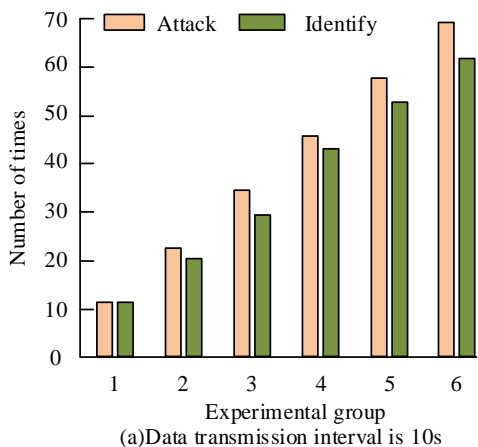


Fig. 11. Attack test results.

V. CONCLUSION

IoT's future development trajectory is to merge with AI as both IoT and AI technology advance. The study suggests a decentralized AIIoTS design based on the BPNN model to enhance the quality of service of IoT devices by leveraging the concept of cloud-fog combination to develop the functionalities of the devices and causing the DC to be in the system to realize the mapping of BPNN and IoTS. The experimental data revealed an average transmission rate of 8.265 Mbps between the sensing device and the FD, which was sufficient to satisfy customer demand. The average communication connection bandwidth between a typical sensor and a cloud server was 3.2 Mbps, whereas the decentralized AIIoTS based on the BPNN model had an average communication link bandwidth of 8.01 Mbps, which increased data transmission efficiency by 4.81 Mbps and increased stability. Users' network data transmission rates were more than 8.5Mbps when the data transmission interval was set to 20s, and as data transmission increased, these rates did not vary substantially and could accommodate users' usage needs. The network data transmission error increased gradually as the number of users increased, but all mistakes were less than 0.07, and the proper network transmission rate was higher than 0.93 at that point, guaranteeing the correctness of user data transfer. When the data transmission interval was 20s, the network was able to recognize 100% of threat attacks, showing that the decentralized AIIoTS based on the BPNN model had stronger data transmission security and better network device service quality. The decentralized artificial intelligence IoT system proposed in the study currently only achieves basic functions, and further improvement is needed to make the system more fully functional. In terms of system stability, the studied system uses network edge devices, and the stability of the system may be relatively poor. For example, if the sensing device, fog device, or execution device may be damaged, how should the system continue to operate correctly? Future research can develop towards the direction of intelligent systems, upgrading devices in the system to intelligent agents, achieving automatic redeployment in the event of system failures, and further developing and intelligentising Docker containers.

REFERENCES

- [1] Li X, Wu J. Node-oriented secure data transmission algorithm based on IoT system in social networks. *IEEE Communications Letters*, 2020, 24(12): 2898-2902.
- [2] Meng A, Gao X, Zhao Y, Yang Z. Three-dimensional trajectory optimization for energy-constrained UAV-enabled IoT system in probabilistic LoS channel. *IEEE Internet of Things Journal*, 2021, 9(2): 1109-1121.
- [3] Yugank H K, Sharma R, Gupta S H. An approach to analyse energy consumption of an IoT system. *International Journal of Information Technology*, 2022, 14(5): 2549-2558.
- [4] Ioannou I, Vassiliou V, Christophorou C, Pitsillides A. Distributed artificial intelligence solution for D2D communication in 5G networks. *IEEE Systems Journal*, 2020, 14(3): 4232-4241.
- [5] Zhao X. Implementation of English ICAI MOOC system based on BP neural network. *Journal of Ambient Intelligence and Humanized Computing*, 2023, 14(4): 3177-3186.
- [6] Xiao M, Wen K, Yang G, Lu X. Research on gearbox fault diagnosis system based on BP neural network optimized by particle swarm optimization. *Journal of Computational Methods in Sciences and Engineering*, 2020, 20(1): 53-64.
- [7] Yang Y H, Shi Y. Application of improved BP neural network in information fusion Kalman filter. *Circuits, Systems, and Signal Processing*, 2020, 39(10): 4890-4902.
- [8] Cheng P, Chen D, Wang J. Clustering of the body shape of the adult male by using principal component analysis and genetic algorithm-BP neural network. *Soft Computing*, 2020, 24(17): 13219-13237.
- [9] Xie X T. Technology enterprise value assessment based on BP neural network. *International Journal of Computing Science and Mathematics*, 2020, 12(2): 192-203.
- [10] Shi J, Chengchao S, Lei H, Mengxi X. Smart grid short-term load estimation model based on BP neural network. *International Journal of Computing Science and Mathematics*, 2020, 11(2): 123-136.
- [11] Deng S, Zhao H, Fang W, Yin J, Dustdar S, Zomaya A Y. Edge intelligence: The confluence of edge computing and artificial intelligence. *IEEE Internet of Things Journal*, 2020, 7(8): 7457-7469.
- [12] Chandrakar M, Patle V K. Security issue in iot based architecture for health care system. *Research Journal of Engineering and Technology*, 2020, 11(2): 89-97.
- [13] Shapsough S Y, Zualkernan I A. A generic IoT architecture for ubiquitous context-aware learning. *IEEE Transactions on Learning Technologies*, 2020, 13(3): 449-464.
- [14] Sarrigiannis I, Ramantas K, Kartsakli E, Mekikis P V, Antonopoulos A, Verikoukis C. Online VNF lifecycle management in an MEC-enabled 5G IoT architecture. *IEEE Internet of Things Journal*, 2019, 7(5): 4183-4194.
- [15] Cheng C F, Chen Y C, Lin C W. A Carrier-Based Sensor Deployment Algorithm for Perception Layer in the IoT Architecture. *IEEE Sensors Journal*, 2020, 20(17): 10295-10305.
- [16] Sun G, Zeng G, Hu C, Jiang T. Starch-based aerogel prepared by freeze-drying: establishing a BP neural network prediction model. *Iranian Polymer Journal*, 2023, 32(1): 37-44.
- [17] Zhou W, Li M, Li L, Yeh K. Prediction of SET on SRAM Based on WOA-BP Neural Network. *Journal of Internet Technology*, 2023, 24(2): 267-273.
- [18] Li T, Deng J, Ren J. Security and Energy Efficiency: Breaking the Barriers of High Peak-to-Average Power Ratio and Disguised Jamming in NextG IoT System Design. *IEEE Internet of Things Journal*, 2022, 10(3): 2658-2666.
- [19] Qi P, Zhou X, Ding Y, Zhang Z, Zheng S, Li Z. Fedbkd: Heterogenous federated learning via bidirectional knowledge distillation for modulation classification in iot-edge system. *IEEE Journal of Selected Topics in Signal Processing*, 2022, 17(1): 189-204.
- [20] Bashar D A. Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence and Capsule Networks*, 2019, 1(2): 73-82.
- [21] Wang X, Cheng M, Eaton J, Hsieh C J, Wu S F. Fake node attacks on graph convolutional networks. *Journal of Computational and Cognitive Engineering*, 2022, 1(4): 165-173.
- [22] Nimrah S, Saifullah S. Context-Free Word Importance Scores for Attacking Neural Networks. *Journal of Computational and Cognitive Engineering*, 2022, 1(4): 187-192.
- [23] Chen Z. Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. *Journal of Computational and Cognitive Engineering*, 2022, 1(3): 103-108.
- [24] Andavan M T, Vairaperumal N. Cloud computing based deduplication using high-performance grade byte check and fuzzy search technique. *Journal of Intelligent & Fuzzy Systems*, 2023, 44(2): 1-15.
- [25] Mohanaprakash T A, Nirmalrani D V. Exploration of various viewpoints in cloud computing security threats. *Journal of Theoretical and Applied Information Technology*, 2021, 99(5): 1172-1183.

Black Widow Optimization Algorithm for Virtual Machines Migration in the Cloud Environments

Chuang Zhou^{1*}

Anhui Vocational College of Defense Technology
Lu'an, Anhui, 237011, China

Abstract—Cloud data centers use virtualization technology to manage computing resources. Using a group of connected Virtual Machines (VMs), corresponding users can compute data efficiently and effectively. It improves the utilization of resources, thereby reducing hardware requirements. Repossession of affected services requires VM-based infrastructure overhaul schemes. Clarifications concerning dedicated routing are also desirable to improve the reliability of Domain Controller (DC) services. The migration of a VM experiencing a node failure challenges maintaining reliability. The selection of VMs is influential in limiting the number of VM migrations. Choosing one or more potential VMs for migration reduces the servers' workload. This paper presents an energy-aware VM migration method for cloud computing based on the Black Widow Optimization (BWO) algorithm. The proposed algorithm was implemented and measured using JAVA. Afterward, we compared our results against existing methodologies regarding resource availability, energy consumption, load, and migration cost.

Keywords—Cloud computing; migration; energy consumption; optimization; black widow algorithm

I. INTRODUCTION

Cloud computing is rapidly consolidating itself as a new computing technology [1]. It aims to provide services and meet user demands for high reliability, scalability, and availability [2]. Cloud computing has gained popularity over the past few years due to its efficient utilization of resources and convenience in accessing services [3]. These competitive advantages can be attributed to virtual technology and distributed networking in the cloud [4]. Storage, processing, memory, bandwidth, network, and virtual machines are some of the resources provided by cloud computing [5]. The provided virtualized services can be categorized as Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS) [6]. Virtualization facilitates access to resources while hiding their physical characteristics [7]. This technique allows several different environments to be separated or shared so that they can interact without being aware of one another [8]. As a matter of fact, virtualization is the key technology behind cloud computing. Data centers typically consist of hundreds of heterogeneous servers that consume significant energy. Increasing emissions and climate change have prompted governments, organizations, and IT enterprises to try and manage data centers more sustainably [9]. Consequently, data centers in the cloud require considerable energy as cloud services grow rapidly [10]. Cloud service providers face a serious challenge in reducing energy

consumption. Two factors cause this situation. The first is that data centers consume a lot of energy, resulting in carbon dioxide emissions that are incompatible with the environment. The second reason is the low power efficiency [11].

The concept of migrating Virtual Machines (VMs) dynamically and transparently from one host to another is called VM migration [12]. In addition, VM migration provides an opportunity to identify hotspots in data centers. In the realm of virtual machine migration within cloud environments, the convergence of IoT, Artificial Intelligence (AI), and Machine Learning (ML) introduces transformative capabilities. IoT sensors enable real-time data collection, facilitating dynamic workload monitoring and resource utilization assessment [13-15]. AI algorithms harness this influx of data to make intelligent decisions on VM migrations, optimizing resource allocation and energy efficiency. ML algorithms, fueled by historical and real-time data, identify patterns and predict workload trends, enabling proactive migration strategies to prevent bottlenecks and ensure optimal performance [16, 17]. The synergy of IoT, AI, and ML empowers cloud systems to autonomously adapt to fluctuating workloads, enhance resource provisioning, and optimize VM migrations, thereby elevating the efficiency and resilience of cloud environments.

Workload regulation on individual nodes via VM migration to optimize energy consumption is a Non-deterministic Polynomial (NP)-hard problem, commonly solved with heuristic approaches. In these cases, metaheuristic algorithms can be used efficiently when local heuristics are insufficient for finding optimum solutions [18]. Metaheuristics are repetitive procedures used as guides and amendments to other heuristics. Based on the Black Widow Optimization (BWO) algorithm, the current study proposes a new VM migration method that reduces energy consumption, hosts, and migrations. The BWO has a short computation time for globally optimized results, a high convergence speed as the quality parameters can be regulated, and a flexible optimization pattern [19]. This algorithm draws inspiration from the distinctive mating conduct observed in black widow spiders. Owing to its unique operators, it can be deemed a fusion of evolutionary and swarm methodologies. Notably, the BWO algorithm introduces a distinctive phase called "cannibalism." This phase holds a significant advantage as it eliminates species exhibiting inadequate fitness from the population, thus fostering an accelerated convergence rate. This paper contributes to the following:

- Multi-objective VM migration model designed for choosing the most suitable VM.
- The BWO algorithm is proposed to select the most efficient VMs in the cloud.

The rest of the paper is organized as follows. The literature related to this topic is briefly reviewed in Section II. Section III describes the proposed algorithm. The experiments and results are summarized in Section IV. Section V outlines the conclusions derived from the analysis.

II. RELATED WORK

Kansal and Chana [20] have proposed a method for migrating VMs based on Firefly Optimization (FFO). The FFO algorithm optimizes effective energy simultaneously at the memory and processor levels. Furthermore, it reduces the number of PMs and VMs, which avoids further energy wastage. VMs with the highest load are transmitted to lower-load nodes to maintain efficiency and performance. Wang, et al. [21] introduced enough green energy-efficient data centers. Diverse renewable energy supplies have been considered in the efficient management of VM migration. The proposed method can flexibly manage green energy and cooling power consumption. The authors assessed the effect of temperature on the energy consumption of cooling and IT devices.

Xu and Abnoosian [12] presented a hybrid optimization algorithm based on genetic and particle swarm optimization algorithms for improving VM energy consumption and execution time during VM migration. In the hybrid algorithm, GA is utilized to overcome the limitations of the PSO algorithm, which suffers from slow convergence and limited global optimization. According to the results, the proposed method has improved energy consumption by an average of 23.19% compared to the other three methods. Results also revealed a 29.01% improvement in execution time over the other three methods. Zhou, et al. [22] introduce an energy-efficient algorithm for VM migrations. This algorithm optimizes host location, VM selection, and trigger time when memory and CPU factors are considered. It migrates some VMs from lightly loaded to heavily loaded hosts using virtualization technology. Energy is conserved by switching idle hosts to the low-power mode or shutting them down. This algorithm reduces SLA violations by 13% and saves 7% of energy over the Double Threshold (DT) algorithm.

Fu, et al. [23] presented a layered VM migration algorithm. Cloud data centers are divided into several regions based on bandwidth utilization rates. VM migrations balance network load between regions, resulting in load balancing of cloud resources. Experiments indicate that the proposed algorithm is shown to be able to balance network resource load in cloud computing effectively. Chien, et al. [24] have proposed an efficient VM migration algorithm based on minimizing migrations in cloud computing to improve efficiency, meet user requirements, and prevent service level agreements (SLA) violations. The proposed algorithm was more effective than existing algorithms based on experimental results. A threshold algorithm was proposed by Kaur and Sachdeva [25] to allocate tasks to the most capable machine and host and to maintain checkpoints on VMs. Overloaded VMs need to migrate tasks to

another VM. This study proposes a weight-based technique for migrating cloudlets between VMs.

Cao and Hou [26] have introduced a bi-level VM placement algorithm. The initial tier incorporates a queuing model devised to manage a multitude of VM requests. This model facilitates the seamless implementation and validation of diverse models, including cloud simulations. Additionally, it furnishes an alternative mechanism for task allocation to servers. Subsequently, a multi-objective VM placement algorithm is introduced, rooted in the Krill Herd (KH) algorithm. Fundamentally, this algorithm strives to strike an equilibrium between energy consumption and the efficient utilization of resources.

Khan and Santhosh [27] have proposed a hybrid optimization algorithm for managing VM migration within a cloud environment. This novel approach amalgamates Particle Swarm Optimization (PSO) and Cuckoo Search (CS) algorithms to yield the proposed hybrid optimization framework. The focal point of this research endeavor is mitigating energy consumption, computation time, and migration expenses. Additionally, a secondary objective pertains to the maximization of resource utilization. To substantiate the research objectives, the efficacy of the hybrid optimization model is rigorously assessed through simulation analysis. This assessment encompasses a comparative study against conventional algorithms. The evaluation parameters encompass computation time, resource availability, migration cost, and energy consumption.

Kumar and Sivakumar [28] have introduced a novel approach for VM migration, hinging on the CS algorithm. The selection of an appropriate provider is undertaken through a comprehensive consideration of multiple constraints, including factors such as delay, bandwidth, cost, and load. Subsequent to this, effective search criteria are calculated, facilitating the identification of the optimal service contingent upon fitness constraints. These search criteria are framed as optimization problems, and their resolution is undertaken using the CS algorithm. The proposed CS algorithm is meticulously designed by integrating the Cuckoo Search Optimization (CSO) technique with the Salp Swarm Algorithm (SSA). This amalgamation ensures the evaluation of the fitness function for optimal VM migration, incorporating diverse parameters encompassing delay, cost, bandwidth, and load. Consequently, the cloud manager can adeptly execute VM migration in the cloud environment, leveraging the proposed CS-based VM migration approach. The performance evaluation of the CS-based VM migration technique focuses on delay, cost, and load. The results demonstrate that the proposed CS-based VM migration methodology achieves commendable outcomes, manifesting in a minimal delay of 0.14, cost of 0.05, and load of 0.18.

These studies address various facets of the problem, each contributing unique insights and methodologies. Notably, challenges related to optimizing energy efficiency, load balancing, resource utilization, and cost reduction are prevalent across these studies. Kansal and Chana [20] focus on optimizing energy efficiency and reducing the number of PMs and VMs. Wang, et al. [21] explore the integration of diverse

renewable energy supplies and cooling power consumption management. Xu and Abnoosian [12] tackle slow convergence and limited global optimization issues in hybrid optimization. Zhou, et al. [17] optimize host location, VM selection, and trigger time for energy-efficient migrations. Fu, et al. [23] propose a layered VM migration algorithm for network load balancing. Chien, et al. [24] emphasize minimizing migrations to prevent SLA violations. Kaur and Sachdeva [25] propose a threshold algorithm for efficient task allocation. Cao and Hou [26] address energy consumption and resource utilization equilibrium. Khan and Santhosh [27] and Kumar and Sivakumar [28] introduce hybrid optimization techniques to mitigate energy consumption and migration costs, and enhance resource utilization while considering multiple constraints. These studies underscore the multifaceted nature of VM migration optimization and lay the groundwork for further advancements in this evolving field.

III. PROPOSED METHOD

In this section, we discuss virtual machine migration using the BWO. The cloud migration model is illustrated in Fig. 1. Cloud models involve multiple PMs for handling user requests, and PMs gather the VMs to perform tasks on demand. The VMs are created dynamically to alleviate the bottlenecks in cloud computing, and virtualization enhances speed. Cloud services are delivered as tasks to users and are assigned to VMs in a round-robin fashion. In this case, PM controls the set of VMs, and a load balancer monitors loads of PMs. VM migration occurs when a load of a PM exceeds a threshold level. The analytical approaches employed in the development of the VM migration model are as follows.

- A cloud with m PMs and n VMs is created initially.
- This stage sets the migration cost of PMs to the highest value, so it equals 1.
- Round-robin assignment is used to assign incoming tasks to VM at the time interval.
- When the VM's load value exceeds the threshold, migrate the VM in an optimal manner using the proposed BWO.
- This step determines qualitative criteria, such as energy consumption, resource availability, and migration costs.
- The algorithm repeats steps (3) to (5) for every iteration, and then it ends.

A. Initialization

Our cloud system consists of n VMs and m PMs. According to Eq. 1, C , PM_1 , and PM_m stand for cloud, the first PM, and m^{th} PM, respectively.

$$C = \{PM_1, PM_2, \dots, PM_m\} \quad (1)$$

Eq. 2 illustrates VMs, with VM_1 being the first one and VM_n being the last one.

$$PM_m = \{VM_1, VM_2, \dots, VM_n\} \quad (2)$$

The cloud is populated by t users involving k tasks. Eq. 3 can be used to indicate each user.

$$U_t = \{T_1, T_2, \dots, T_k\} \quad (3)$$

In addition, a round-robin process assigns the users' tasks to VMs. In the cloud model, VMs are determined by several parameters, such as memory, bandwidth, CPU, and processing power. VMs in cloud computing environments have the following characteristics, as defined by Eq. 4.

$$V_m^n = \{I_m^n, D_m^n, B_m^n, C_m^n, J_m^n\} \quad (4)$$

I_m^n denotes the total number of utilized MIPS, D_m^n refers to memory, B_m^n signifies the bandwidth, C_m^n represents the number of CPUs, and J_m^n stands for the total number of processing entities. A scale of 1 to 10 is given for the above parameters.

B. Load computation

The load is calculated based on the resources needed by VMs to process tasks supplied by the user. A cloud load is evaluated by considering the processing power, CPU, memory, MIPS, and bandwidth using Eq. 5, in which t represents time, and RU stands for resource utilization. Eq. 6 calculates the resource used by m^{th} VM present. The variables in Eq. 6 are defined in Table I.

$$Load(L) = \frac{R_U}{t} \quad (5)$$

$$R_U = \frac{1}{F} \sum_{i=1}^M \left(\frac{I^F}{\max(I^F)} + \frac{D^F}{\max(D^F)} + \frac{B^F}{\max(B^F)} + \frac{C^F}{\max(C^F)} + \frac{J^F}{\max(J^F)} \right) \quad (6)$$

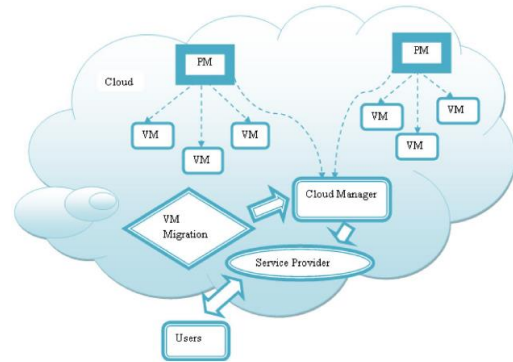


Fig. 1. VM migration model.

TABLE I. VARIABLES IN EQ. 6.

Variable	Definition
F	Normalization factor
m	Number of VMs in each PM
I^F	MIPS
D^F	Memory
B^F	Bandwidth
C^F	CPU
J^F	Processing entity

C. Resource Availability

It is responsible for ensuring the efficient use of resources. Load balancing must be optimized for well-organized performance. Eq. 7 calculates resource availability.

$$R_A = 1 - R_U \quad (7)$$

D. Migration Cost

A VM migration cost is determined by the number of movements. Migration costs for the entire cloud environment can be calculated by Eq. 8, in which c stands for constant, M refers to the total number of VMs, G denotes the number of migrations, and N is the total number of PMs.

$$M_c = \frac{1}{N} \sum_{j=1}^n \left(\frac{G}{c \times M} \right) \quad (8)$$

E. Energy Model

Each VM consumes energy to migrate and process data. In this way, cloud setup energy is primarily dependent on the power consumed by the resources within the VM, calculated by Eq. 9, in which T represents the total duration, and K denotes the power consumed by the VM calculated by Eq. 10. The maximum power consumed is denoted by K_{max} , and the resource utilization is represented by R_U^{cloud} calculated by Eq. 11.

$$E = \frac{1}{T} \sum_{t=1}^T K \quad (9)$$

$$K = p \times K_{max} + (1 - p) \times K_{max} \times R_U^{cloud}; \quad 1 < p < t < T \quad (10)$$

$$R_U^{cloud} = \frac{1}{N} \sum_{r=1}^N R_U \quad (11)$$

F. Solution encoding

The proposed algorithm finds suitable VMs for migration. Consider a scenario in which PM 1 involves 2 VMs, and PM 2 involves 3 VMs. A round-robin assignment process assigns the incoming tasks to VMs. All processes are handled in a circular order, according to assigned time slices, with no priority given to any process. VMs are migrated to under-loaded PMs if they exceed a threshold value, and an optimization algorithm is used to assign the optimal VM. A solution encoding for identifying optimal VMs for migration is shown in Fig. 2.

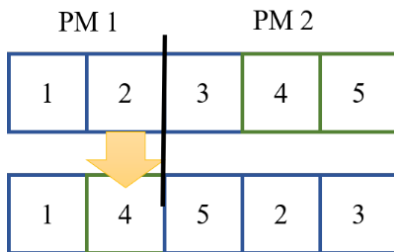


Fig. 2. Encoding solutions.

G. Fitness function

An optimal solution is determined by computing the fitness function. Energy, migration cost, load, and resource utilization are the parameters that are used in formulating the fitness function. VMs are selected based on the fitness function, calculated by Eq. 12.

$$Fitness(f) = \delta(1 - E) + \gamma(1 - M_c) + \beta(1 - L) + \alpha R_A \quad (12)$$

In Eq. 12, E stands for energy, M_c represents migration costs, L refers to the load, R_A denotes resource utilization, δ , γ , β , and α are the weights ranging from 0 to 1.

H. BWO algorithm for VM migration

This section applies the BWO algorithm to the multi-objective VM migration method. VMs are migrated to specific PMs based on cost, energy, and load. The performance of the VM migration algorithm is determined by the reduction in energy consumption, cost, and load. As with traditional methods, BWO starts by initializing a population of spiders representing possible solutions. New generations of spiders are produced in pairs. As part of this optimization, female spiders eat black male spiders during or after mating. Female black widow spiders release their stored sperm into egg sacs after storing them in their sperm theca. The spiderling emerges from its egg sac after 11 days. A spiderling remains in its mother's web for several days to a week, during which sibling cannibalism occurs. When the wind blows, the spiderling leaves the web.

- Initializing the population

Initially, the population is defined by the number of spiders, each representing a possible solution. An individual black widow represents a solution to an optimization problem in d -dimensional space. Eq. 13 defines the array.

$$Black\ widow = [x_1, x_2, \dots, x_d] \quad (13)$$

The array contains floating point variables. There are NP black widow spiders in a black widow population, expressed as an $NP \times d$ candidate matrix. Eq. 14 assigns the initial population at random.

$$Black\ widow = xl + rand(l, d) \times (xu - xl) \quad (14)$$

- Procreate

As each pair is independent of the other, it starts mating to reproduce a new generation, just as each couple naturally mates in its web, independently of the rest of the web. Even though hundreds of eggs are laid every time a spider mates, the number of muscular spider babies remains the same. The BWO produces offspring by reproducing an array called alpha using arbitrary numbers, in which u_1 and u_2 represent parents, whereas v_1 and v_2 represent children.

$$\begin{cases} v_1 = a \times u_1 + (1 - a) \times u_2 \\ v_2 = a \times u_2 + (1 - a) \times u_1 \end{cases} \quad (15)$$

- Cannibalism

Three categories of cannibalism are included in this algorithm. One of the earliest forms is cannibalism, where a

female black widow eats a black male widow after or during mating. Fitness values are used to identify female and male black widows. Sibling cannibalism is another type of cannibalism in which stronger siblings eat weaker siblings. The algorithm sets the cannibalism rating (CR) based on survivor numbers. The third type of cannibalism occurs when the baby spider eats the mother spider. Weak or strong spiderlings are evaluated based on their fitness value.

- Mutation

Based on the simulated binary crossover (SBC), the BWO algorithm produces new chromes or individuals at constant crossover and mutation rates. A mutation rate is altered using an adaptive scheme, followed by a projection of the enhanced mutation rate. It combines three crossover operators, single-point crossover (SPC), uniform crossover (UC), and SBC, to generate new individuals. As a result, the proposed algorithm is evaluated on three variables. BWO algorithm fails to determine an optimal solution to the optimization problem based on the permanent mutation rate. With the adaptive strategy employed in this paper, the mutation rate can be altered in order to solve this problem. We present a linear function that alters the mutation rate for ease of use. It is, therefore, necessary to update the mutation rate using Eq. 16, in which p_s is calculated by Eq. 17.

$$m_r = \frac{p_s}{L} \tag{16}$$

$$p_s = P + (t - 1) \times \frac{1 - P}{t_M - 1} \tag{17}$$

In Eq. 17, t_M and t stand for the maximum and current generation, respectively, P refers to the fixed real number given by Eq. 18.

$$P_0 = \frac{L}{50} \tag{18}$$

IV. EXPERIMENTAL RESULTS

In this section, we present results obtained from the proposed algorithm and compare them with previous methods. The developed algorithm was tested in JAVA on a PC with Windows 8 and 8GB of RAM. Simulation is conducted in a cloud-based environment containing 10 PMs and 50 VMs, with 25 incoming tasks. The main evaluation indicators are resource availability, energy, migration cost, and load. VM migration techniques on cloud computing platforms commonly use these performance metrics. We compared WOA [29], Firefly [20], and ABC-BA [30] with our proposed algorithm for analysis.

An analysis of 25 incoming tasks is conducted by comparing the developed algorithm with respect to migration cost, resource availability, energy, and load. In Fig. 3, resource availability is compared over a variety of iterations. For the 50th iteration, WOA, Firefly, and ABC-BA have resource availability values of 0.96, 0.952, and 0.938, respectively, which are lower than our algorithm. Fig. 4 illustrates the energy cost analysis. In the 60th iteration, existing techniques, such as WOA, Firefly, and ABC-BA, possess energy costs of 0.498, 0.494, and 0.494, respectively, which are higher than our algorithm. Fig. 5 shows a comparative analysis based on

migration costs. At 60 iterations, WOA, Firefly, ABC-BA, and our algorithm achieved migration cost values of 0.195, 0.132, 0.0605, and 0.059, respectively. Fig. 6 shows the results of the load analysis based on different iterations. At 50 iterations, WOA, Firefly, ABC-BA, and our algorithm have corresponding load values of 0.0098, 0.0038, 0.0029, and 0.0019.

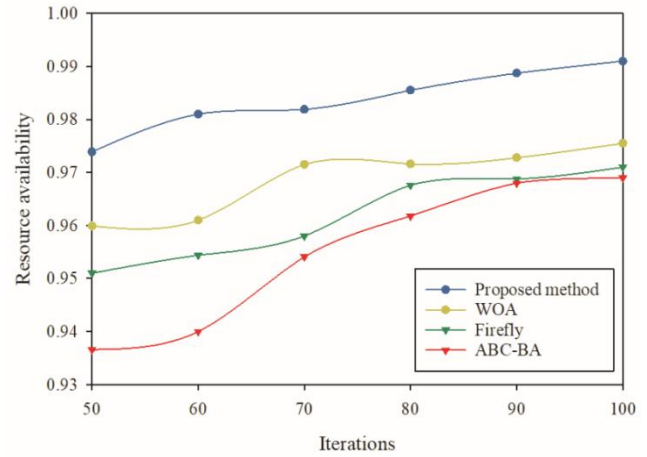


Fig. 3. Resource availability comparison.

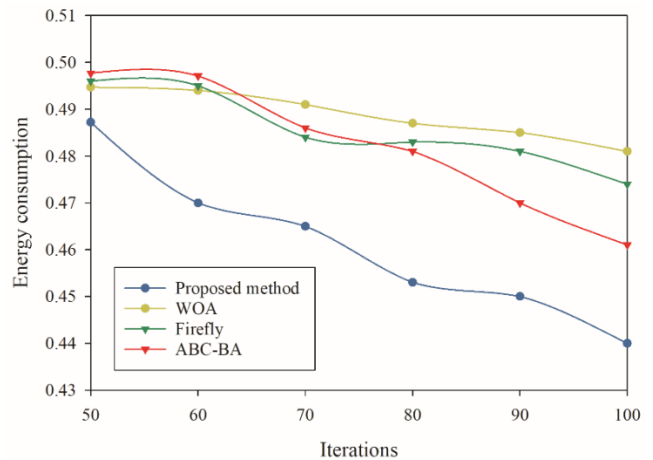


Fig. 4. Energy consumption comparison.

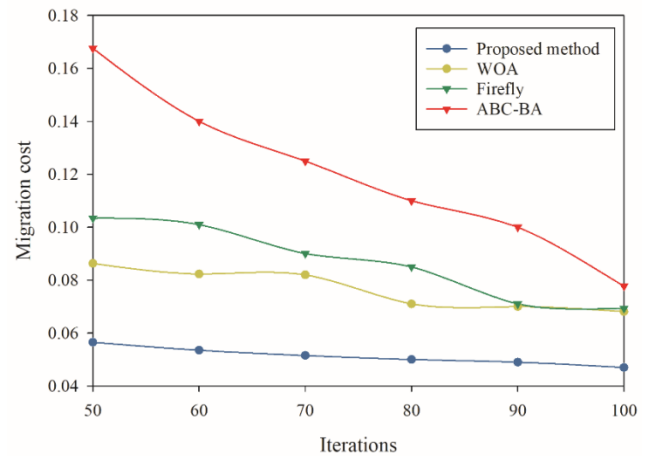


Fig. 5. Migration cost comparison.

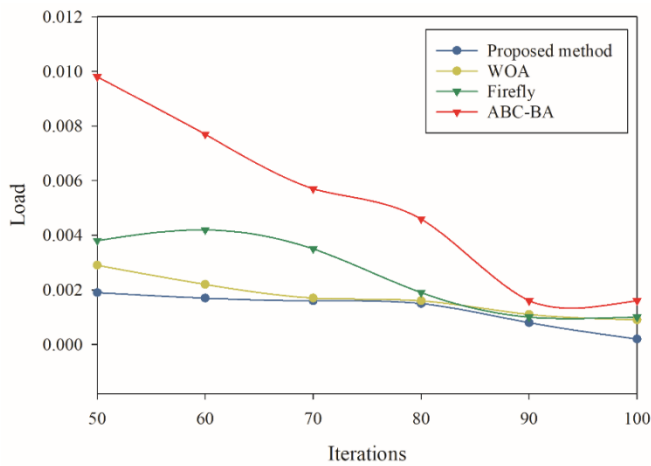


Fig. 6. Load comparison.

The datasets used in our tests were chosen with attention to replicating different characteristics of real-world cloud computing infrastructures. We simulate the complicated dynamics and complexity of existing cloud infrastructures by using datasets with 10 PMs and 50 VMs accommodating 25 incoming tasks. These datasets contain a variety of resource capabilities, imitating real-world cloud setups and emerging as a relevant depiction of dynamic workload management concerns. The intrinsic heterogeneity of these datasets, which includes the interaction of PMs, VMs, and tasks, affects the performance of VM migration algorithms. The datasets' intricacies directly influence resource use, energy consumption, migration cost, and load distribution, all of which are critical performance measures in our study. Because cloud services are resource-intensive, efficient and adaptable migration solutions are required, which our suggested algorithm tries to meet. As we examine the differences in comparing outcomes, it becomes clear that the success of our algorithm stems from its capacity to negotiate the complexities of various datasets intelligently. The algorithm's flexibility in changing situations, such as dynamic workloads and fluctuating resource availability, presents it as a reliable option for optimizing VM migrations.

V. CONCLUSION

Due to the inherent benefits of cloud computing, the number of cloud users and their workloads is increasing daily. Besides, service providers are challenged by maintaining QoS under heavy workloads. Cloud computing can offer better computing services by utilizing VM migration techniques, reducing delays, and optimizing energy usage. This paper successfully developed the BWO algorithm to migrate VMs to the cloud. This algorithm reduces unnecessary migrations by identifying the optimal solution for migrating VMs to the host. Based on simulation results, the proposed algorithm is more efficient regarding migration cost, resource availability, load, and energy consumption than previous methods.

ACKNOWLEDGMENT

Anhui Provincial Education Department Scientific Research Project (KJ2021A1491).

REFERENCES

- [1] B. Pourghebleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, pp. 1-24, 2021.
- [2] S. Sefati, M. Mousavinasab, and R. Zareh Farkhady, "Load balancing in cloud computing environment using the Grey wolf optimization algorithm based on the reliability: performance evaluation," *The Journal of Supercomputing*, vol. 78, no. 1, pp. 18-42, 2022.
- [3] A. Najafizadeh, A. Salajegheh, A. M. Rahmani, and A. Sahafi, "Multi-objective Task Scheduling in cloud-fog computing using goal programming approach," *Cluster Computing*, vol. 25, no. 1, pp. 141-165, 2022.
- [4] H. Jin, S. Lv, Z. Yang, and Y. Liu, "Eagle strategy using uniform mutation and modified whale optimization algorithm for QoS-aware cloud service composition," *Applied Soft Computing*, vol. 114, p. 108053, 2022.
- [5] M. R. Dorsala, V. Sastry, and S. Chapram, "Blockchain-based solutions for cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 196, p. 103246, 2021.
- [6] J. Yang, B. Jiang, Z. Lv, and K.-K. R. Choo, "A task scheduling algorithm considering game theory designed for energy management in cloud computing," *Future Generation computer systems*, vol. 105, pp. 985-992, 2020.
- [7] Y. Wang, J. Wen, Q. Wu, L. Guo, and B. Tao, "A dynamic cloud service selection model based on trust and SLA in cloud computing," *International Journal of Grid and Utility Computing*, vol. 10, no. 4, pp. 334-343, 2019.
- [8] X. Wei, "Task scheduling optimization strategy using improved ant colony optimization algorithm in cloud computing," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-12, 2020.
- [9] Y. Karaca, M. Moonis, Y.-D. Zhang, and C. Gezgez, "Mobile cloud computing based stroke healthcare system," *International Journal of Information Management*, vol. 45, pp. 250-261, 2019.
- [10] A. Rezaeiapanah, M. Mojarad, and A. Fakhari, "Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic," *International Journal of Computers and Applications*, vol. 44, no. 2, pp. 139-147, 2022.
- [11] G. J. Ibrahim, T. A. Rashid, and M. O. Akinsolu, "An energy efficient service composition mechanism using a hybrid meta-heuristic algorithm in a mobile cloud environment," *Journal of parallel and distributed computing*, vol. 143, pp. 77-87, 2020.
- [12] Y. Xu and K. Abnoosian, "A new metaheuristic - based method for solving the virtual machines migration problem in the green cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 3, p. e6579, 2022.
- [13] R. Singh et al., "Analysis of Network Slicing for Management of 5G Networks Using Machine Learning Techniques," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [14] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23-34, 2017.
- [15] P. He, N. Almasifar, A. Mehbodniya, D. Javaheri, and J. L. Webber, "Towards green smart cities using Internet of Things and optimization algorithms: A systematic and bibliometric review," *Sustainable Computing: Informatics and Systems*, vol. 36, p. 100822, 2022.
- [16] J. Webber, A. Mehbodniya, Y. Hou, K. Yano, and T. Kumagai, "Study on idle slot availability prediction for WLAN using a probabilistic neural network," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*, 2017: IEEE, pp. 1-6.
- [17] M. Bagheri et al., "Data conditioning and forecasting methodology using machine learning on production data for a well pad," in *Offshore Technology Conference*, 2020: OTC, p. D031S037R002.
- [18] R. W. Ahmad, A. Gani, S. H. A. Hamid, M. Shiraz, A. Yousafzai, and F. Xia, "A survey on virtual machine migration and server consolidation frameworks for cloud data centers," *Journal of network and computer applications*, vol. 52, pp. 11-25, 2015.

- [19] V. Hayyolalam and A. A. P. Kazem, "Black widow optimization algorithm: A novel meta-heuristic approach for solving engineering optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 87, p. 103249, 2020.
- [20] N. J. Kansal and I. Chana, "Energy-aware virtual machine migration for cloud computing-a firefly optimization approach," *Journal of Grid Computing*, vol. 14, no. 2, pp. 327-345, 2016.
- [21] X. Wang, Z. Du, Y. Chen, and M. Yang, "A green-aware virtual machine migration strategy for sustainable datacenter powered by renewable energy," *Simulation Modelling Practice and Theory*, vol. 58, pp. 3-14, 2015.
- [22] Z. Zhou, J. Yu, F. Li, and F. Yang, "Virtual machine migration algorithm for energy efficiency optimization in cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 24, p. e4942, 2018.
- [23] X. Fu, J. Chen, S. Deng, J. Wang, and L. Zhang, "Layered virtual machine migration algorithm for network resource balancing in cloud computing," *Frontiers of Computer Science*, vol. 12, no. 1, pp. 75-85, 2018.
- [24] N. K. Chien, V. S. G. Dong, N. H. Son, and H. D. Loc, "An efficient virtual machine migration algorithm based on minimization of migration in cloud computing," in *International Conference on Nature of Computation and Communication*, 2016: Springer, pp. 62-71.
- [25] G. Kaur and R. Sachdeva, "Virtual machine migration approach in cloud computing using genetic algorithm," in *Advances in Information Communication Technology and Computing*: Springer, 2021, pp. 195-204.
- [26] H. Cao and Z. Hou, "Krill Herd Algorithm for Live Virtual Machines Migration in Cloud Environments," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023.
- [27] M. S. A. Khan and R. Santhosh, "Hybrid optimization algorithm for vm migration in cloud computing," *Computers and Electrical Engineering*, vol. 102, p. 108152, 2022.
- [28] A. Kumar and P. Sivakumar, "Cat-squirrel optimization algorithm for VM migration in a cloud computing platform," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1-23, 2022.
- [29] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51-67, 2016.
- [30] K. Karthikeyan et al., "Energy consumption analysis of Virtual Machine migration in cloud using hybrid swarm optimization (ABC-BA)," *The Journal of Supercomputing*, vol. 76, no. 5, pp. 3374-3390, 2020.

Towards Secure Blockchain-enabled Cloud Computing: A Taxonomy of Security Issues and Recent Advances

Shengli LIU*

Department of Public Basic Education, Hebi Polytechnic
Henan Hebi, 458030, China

Abstract—Blockchain technology offers a promising solution for addressing performance and security challenges within distributed systems. This paper presents a comprehensive taxonomy of security issues in cloud computing and explores recent advances in utilizing blockchain to enhance security and efficiency in this domain. We employ a systematic literature review approach to analyze various blockchain-enabled solutions for cloud computing. Our findings reveal that blockchain's decentralized and immutable nature empowers cloud computing services to establish secure and private data interactions. By leveraging blockchain's consensus mechanism, we demonstrate the feasibility of creating a robust platform for authenticating transactions involving digital assets. Through cryptographic methods, blocks of transactions are securely linked, ensuring data integrity. This paper provides a roadmap for understanding security concerns in cloud computing and offers insights into the potential of blockchain technology. We conclude by outlining future research directions that can drive innovation in this exciting intersection of fields.

Keywords—Cloud computing; security; blockchain; review

I. INTRODUCTION

Cloud computing has gained considerable attention in recent years owing to its affordability, sustainability, reliability, scalability, and flexibility. Under a pay-per-use model, it provides on-demand access to infinite virtual resources such as computing, storage, and networks [1]. This scalable and flexible approach to resource delivery has attracted many organizations and individuals. Cloud computing has enabled many enterprises to migrate, compute, and host their applications, giving them seamless access to a range of services without hassle [2]. It is reported that approximately 60 percent of organizations use cloud services to meet their resource needs, accounting for nearly 15 percent of global IT spending [3]. The cloud can efficiently manage bursts of heterogeneous data. It serves as a bridge between end users and the middleware of devices within the IoT architecture [4]. There is a great deal of concern about security in the cloud, which encompasses power consumption, product lifespan, and overall performance [5]. CCTV cameras, social media, and other cloud devices can be accessed and compromised in public places. A Brute Force attack has been conducted on the cloud application due to a weak authentication scheme [6].

The complexity of the cloud computing model and the shared technologies have raised security concerns despite the

obvious benefits [7]. Several elements are involved in the cloud paradigm, such as the network, architecture, APIs, and hardware, which increases the complexity of security issues [8]. This may result in security vulnerabilities if a cloud provider or client uses different configurations. While cloud computing offers organizations benefits such as cost savings, measurable services, rapid scalability, and elasticity, it also introduces inherent risks that must not be overlooked. Cloud computing systems inherently possess various vulnerabilities, giving rise to significant security concerns [9]. Organizations may hesitate to adopt cloud computing despite its potential if they lack robust security policies. An illustration of the advantages of cloud computing can be found in Fig. 1.

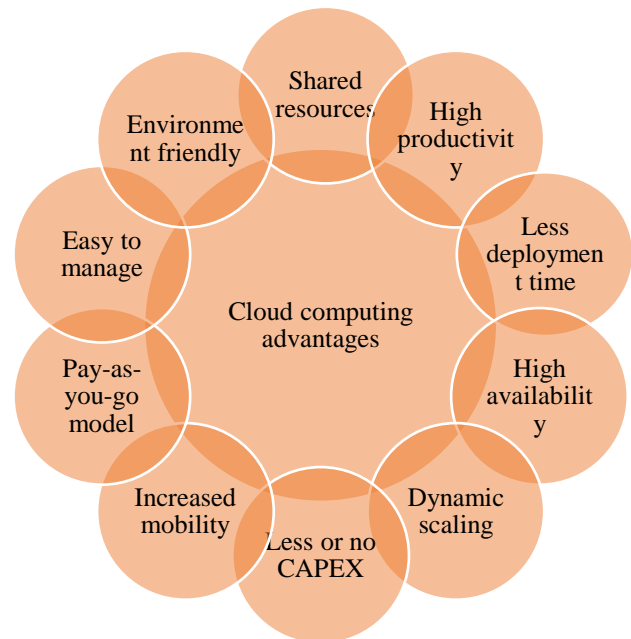


Fig. 1. Cloud computing advantages

Cloud computing offers users on-demand access to customized computing resources, such as services, applications, storage, and servers; instantly delivered by a service provider requiring little management effort. This allows users to access the latest technologies and to scale resources up or down quickly and easily, depending on their needs [10]. It also reduces the costs associated with software and hardware purchases and maintenance. Fig. 2 illustrates four ways cloud

computing can be implemented: public, private, community, and hybrid. In its simplest form, a public cloud is a cloud environment that is publicly accessible by a large number of cloud customers without any restrictions imposed by the cloud provider. Private cloud environments offer the same advantages as public clouds, but access is limited to a specific user or organization. Community clouds provide a shared cloud environment for a specific group of users and organizations. Hybrid clouds combine public and private clouds, offering more flexibility and scalability [11].

Cloud computing encompasses three main service types: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). SaaS allows users to access applications hosted in the cloud, while PaaS offers a platform for developing, running, and managing applications. IaaS, on the other hand, provides the necessary infrastructure, including servers and storage, to run applications. Cloud computing is characterized by key attributes: measured service, rapid elasticity, resource pooling, broad network access, and on-demand self-service [12]. Several metrics can be used to quantify cloud services, including bandwidth, data, and time. Cloud computing services are typically priced according to the number of resources used, and when compared to traditional IT solutions, these services can result in significant cost savings [13]. In cloud computing, the concept of elasticity is used to describe the ability of the system to respond to changes in workloads through automatic provisioning and de-provisioning, as well as the availability of resources. Resource pooling involves pooling virtual and physical resources and allocating and reallocating them dynamically according to consumer demand in a multitenant environment. Broad network access refers to the ability to locate and access resources on a network using various devices and computing platforms, such as tablets, smartphones, laptops, and desktop computers. In the context of on-demand self-service, users have access to their data and resources in the cloud whenever

they need them without requiring assistance from a human [14].

Artificial intelligence (AI) and machine learning (ML) play a pivotal role in the synergy of blockchain-enabled cloud computing, ushering in new frontiers of efficiency and security. AI algorithms leverage the immense volumes of data stored in the blockchain to uncover insights, predict patterns, and optimize resource allocation within cloud systems. ML algorithms enhance consensus mechanisms by dynamically adapting to network demands and mitigating latency [15, 16]. Moreover, AI-driven anomaly detection and threat analysis fortify cloud security by identifying and preemptively mitigating potential breaches. This amalgamation empowers cloud computing with self-optimizing capabilities and real-time threat response, elevating the potential for creating resilient and adaptive cloud ecosystems that harness both the transparency of blockchain and the intelligence of AI and ML [17].

Blockchain technology's potential to address security and performance challenges within distributed systems has garnered significant attention. Integrating blockchain presents a compelling avenue for innovation in the context of cloud computing, which necessitates robust security measures and efficient data management. While existing literature acknowledges the potential of blockchain in enhancing cloud security, a thorough examination of the distinctive security issues and recent advancements in this intersection remains limited. To bridge this gap, our work offers a comprehensive taxonomy of security concerns specifically tailored to blockchain-enabled cloud computing. Beyond traditional gap analyses, we delve into the nuanced intricacies of how blockchain uniquely tackles security and data integrity concerns within the cloud environment. Furthermore, we present an in-depth review of recent advancements that leverage blockchain to reinforce cloud computing's security framework.

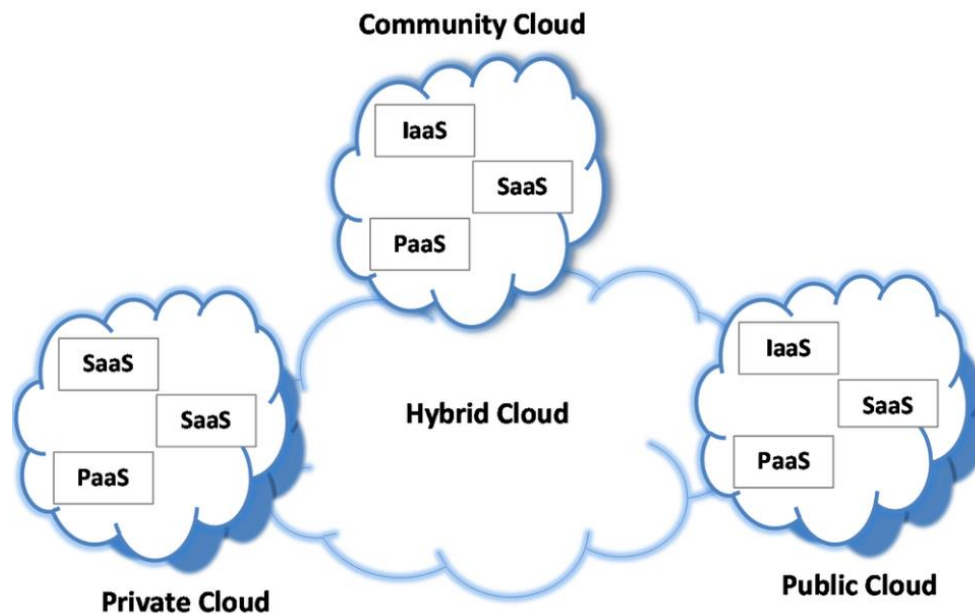


Fig. 2. Cloud deployment models and infrastructure.

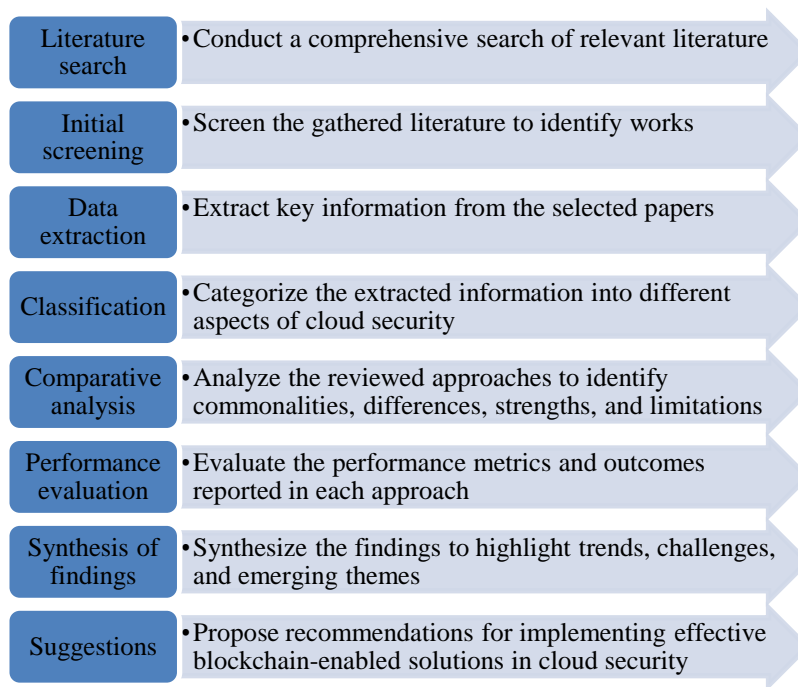


Fig. 3. Review process for blockchain-enabled cloud security approaches.

To systematically review and assess blockchain-enabled cloud security approaches, we have developed a block diagram that illustrates the step-by-step process (See Fig. 3). Beginning with a comprehensive literature search (Step 1), relevant sources are identified and screened (Step 2) to extract key information from selected papers (Step 3). This information is then classified into various aspects of cloud security, such as data integrity and access control (Step 4). A comparative analysis (Step 5) highlights commonalities and differences among the reviewed approaches. Performance evaluation (Step 6) is a crucial element wherein metrics like scalability and efficiency are assessed. The synthesis of findings (Step 7) allows us to outline trends, challenges, and emerging themes in blockchain-enabled cloud security. Moreover, the diagram visualizes the identification of new blockchain solutions (Step 8) that address gaps in existing literature. Lastly, recommendations (Step 9) for the implementation of effective blockchain solutions in cloud security conclude the process. This block diagram visually represents our rigorous approach to evaluating and enhancing cloud security through blockchain integration.

The remainder of the paper is organized in the following manner. Section II delves into a comprehensive review of existing literature on cloud computing security. Section III focuses on our analysis of blockchain technology's potential in enhancing cloud security. Section IV presents our research findings. Section V concludes the paper.

II. LITERATURE REVIEW

Cloud security plays a crucial role in safeguarding all layers of computing in both public and private clouds. As illustrated in Fig. 4, cloud applications benefit from three levels of

protection: SaaS, PaaS, and IaaS. This study focuses on analyzing the existing challenges associated with cloud security and exploring the latest advancements in security solutions. It aims to provide insights into the evolving landscape of cloud security and identify effective measures to mitigate risks and protect cloud-based applications. There are 28 security problems described in the article that can be categorized into five groups (Table I). Comparative evaluations can also be conducted on the latest security technologies and countermeasures. Table I summarizes five types of cloud computing safety concerns. In [10], the same method is used to classify problems, but only for small groups and not for all four types.

Fig. 5 provides an overview of potential security risks associated with different components of the cloud. The cloud infrastructure, clients, and network are all vulnerable to security threats, necessitating the implementation of preventive, detective, and responsive strategies. Table I categorizes these components according to their respective cloud security categories. To enhance security, various security specifications such as SSL, TLS, XML signatures, Interoperability key management protocols, and XML Encryption Syntax and Processing are necessary. Currently, there is a lack of widely accepted security standards specific to cloud computing. While safety requirements may be appropriately defined, compliance risks significantly impact several security issues. The absence of effective governance and evaluation of corporate standards exacerbates this problem. In particular, cloud clients often lack sufficient knowledge regarding the provider's protocols, processes, and activities, particularly in the areas of identity management and job separations.

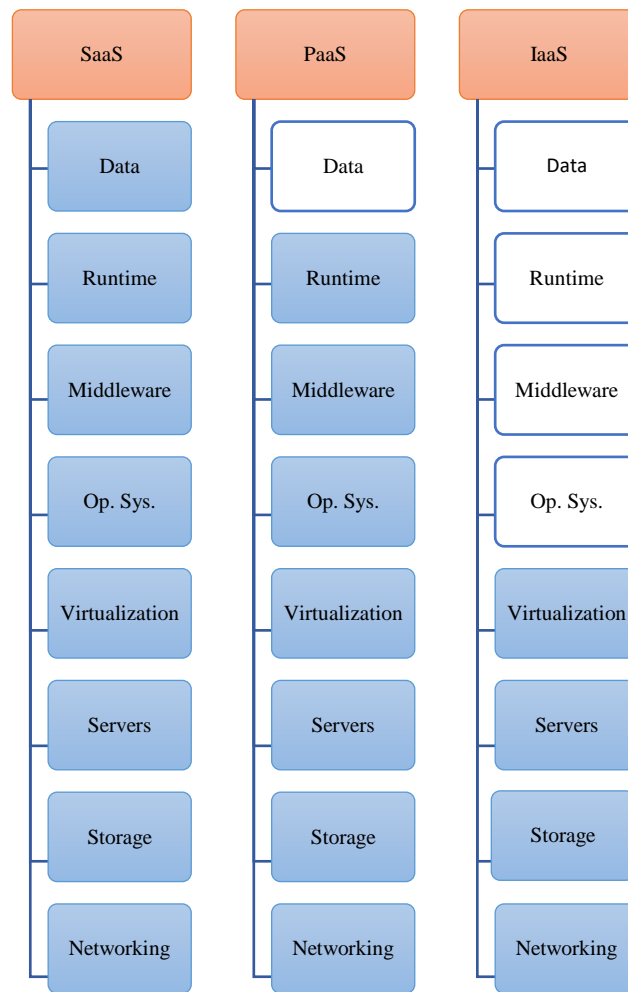


Fig. 4. Cloud security aspects

TABLE I. A TAXONOMY OF CLOUD SECURITY ISSUES

Group	Description	Issues
Data	It addresses data storage, privacy, and data migration issues concerning data security.	Accessibility, protection, privacy, recovery, placement, data loss, and redundancy information
Cloud infrastructure	It focuses on specific threats related to cloud infrastructure	QoS, server location and backup, security misconfiguration, multi-tenancy, reliability of suppliers, and sharing of technical faults
Access control	It is concerned with authentication and connectivity issues and identifies concerns regarding user privacy and data storage.	Browser protection, malicious insider, privileges of the user, and authentication mechanism
Network	It encompasses network intrusions such as link access, DDoS, DoS, flooding attacks, and bugs in the IP protocol.	Installation of the right network firewalls, Internet dependence, IP vulnerabilities, and network security configuration
Security standards	It clarifies the requirements for cloud storage as well as preventive measures to prevent unauthorized access. It specifies cloud computing safety regulations without compromising reliability and efficiency.	Inadequacies in security standards, legal issues, audit failures, enforcement risks, and trust

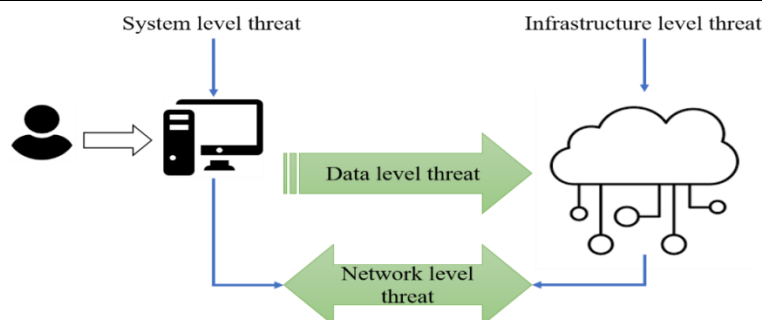


Fig. 5. Security threats associated with cloud components.

The audibility of cloud computing is one of its most critical features. However, there is no network of audit providers for cloud computing services. Ensuring auditability is crucial when a service provider outsources a non-transparent service. It is essential that the entire process is auditable to maintain transparency and accountability. However, security standards and governing bodies that protect Service Level Agreements (SLAs) and regulatory matters are not currently incorporated into cloud computing practices. This absence of established security standards and governing bodies poses challenges in effectively addressing and mitigating risks associated with outsourced services in cloud computing environments. Considering the vulnerability of cloud computing to network-related security attacks, network problems pose the greatest threat to cloud security. Operations in the cloud are heavily dependent on networking and are closely interconnected. The quality of service (QoS) has emerged as an unexpected challenge in the cloud computing landscape, as numerous service providers aim to offer fast and cost-effective performance. We consider QoS to be a critical factor that directly or indirectly influences security. Even a minor error in the configuration of one or more cloud components can have a profound impact on multiple services, considering that cloud configurations are often shared among numerous services. Various case studies emphasize the significance of encrypting, securing, managing, and ensuring timely access to data. This highlights the importance of addressing QoS concerns to ensure the overall security and effectiveness of cloud computing services. Several issues have been identified as major concerns in the literature, including data availability [18], data security [19], data confidentiality [20], data recovery [21], data localization [22], data loss [9], and data redundancy [23].

Blockchain technology refers to a network of blocks containing user records safeguarded using cryptography. These blocks are interconnected, allowing for the distribution of information throughout the network. The concept of blockchain was initially introduced in 1991 by Stuart Haber and W Scott Stornetta [24]. It was later implemented by an anonymous developer known as Satoshi Nakamoto, who used it in the creation of the digital currency Bitcoin. Initially designed for Bitcoin, blockchain technology has now found applications in various domains. Researchers have explored its potential in securing financial transactions, contracts, inter-organizational transactions, IoT systems, banking, land records, and more. Bitcoin's success in maintaining distributed ledgers and transactions without the involvement of a central authority has been instrumental in advancing blockchain technology. While Bitcoin continues to operate on Nakamoto's original blockchain, other projects like Ethereum and Ripple have emerged, each with its own set of rules and regulations and a wider range of applications [25].

Decentralization, transparency, and immutability are three characteristics of blockchain. Decentralization entails that the blockchain distributes its contents, which means the blockchain does not have a single owner. Transparency indicates that transaction information can be viewed only by a user's public address. A blockchain cannot be modified due to its immutability. These characteristics create an immutable and

secure network resistant to hacking and tampering. The transactions stored on the blockchain cannot be reversed and are fully traceable. The distributed nature of the blockchain also allows data to be stored securely and reliably [26].

Fig. 6 provides an illustration of the structure of blockchain and its associated technologies. A block is composed of two main components: the header and the transactions. The header contains the hash value of the previous block and a unique nonce number. The transactions part contains information about all the transactions included in the block. Each block includes the hash value of the previous block, which is a combination of the previous block's hash value and the current block's hash value. This property ensures the immutability of the blockchain. If an attacker attempts to modify the hash value by even a single bit, it will result in a change in the hash value of the subsequent block. This ripple effect continues throughout the entire blockchain. The attacker would need to recalculate the hash value of all the following blocks, which is extremely challenging.

Blockchain technology enables various applications in fields such as healthcare, finance, distribution, and more. One of the key features of blockchain is its ability to facilitate peer-to-peer transactions without the need for a centralized authority. The foundation of blockchain technology is built upon principles of trust and security, which are enabled through cryptography. By establishing peer-to-peer communication within a decentralized network, blockchain technology allows for trust to be established among unknown peers. The use of public and private keys plays a crucial role in ensuring security within the blockchain. A public key serves as a shared address known to everyone in the network, similar to an email address. On the other hand, a private key is a unique address that is only accessible to the user, similar to an email password.

Software programmers play a crucial role in verifying and validating transactions on the blockchain. To enhance and streamline transactions, innovative technologies have been incorporated into the computational elements of the blockchain. These transactions are recorded in a distributed ledger to ensure transparency and immutability. While blockchain technology is integral to modern digital systems, it does have certain limitations. The small size of blocks restricts the number of transactions that can occur within the network, leading to longer block creation times and reduced throughput. Fig. 7 provides an overview of the transactional flow in blockchain technology. Nodes serve as the foundation of blockchain architecture, with users or highly configured computers acting as nodes. Each node maintains a complete copy of the blockchain ledger. Miners, which are specialized nodes, have the capability to add new blocks to the blockchain. Users undergo authentication, verification, and validation processes by miners. Once a transaction is authenticated and validated by miners, the corresponding amount is deducted from the sender's wallet and credited to the receiver's wallet. The concept of a block can be likened to a container that holds an aggregated set of transaction details. New transactions initiated on the blockchain result in the creation of a new block, which can only be added to the blockchain after successful verification by miners.

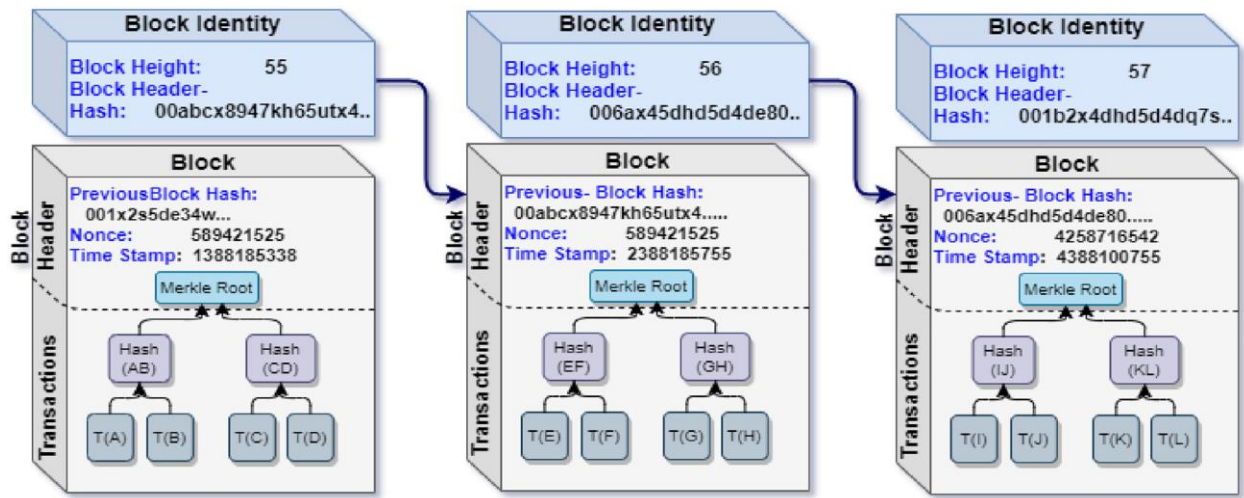


Fig. 6. Blockchain structure.

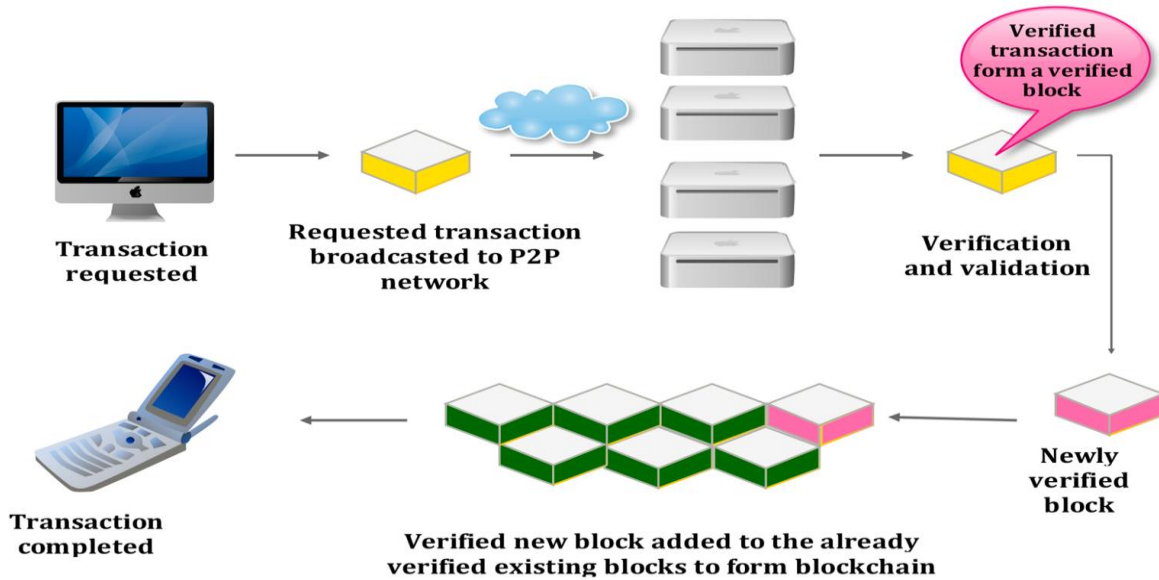


Fig. 7. Transaction flow blockchain technology.

III. BLOCKCHAIN-ENABLED CLOUD SECURITY APPROACHES

Blockchain technology and smart contracts have impacted a wide range of engineering and computer science disciplines. Cloud computing can benefit from blockchain technology by re-engineering the data centers. Due to its decentralized operating model, blockchain can potentially replace centralized cloud-based services. Blockchain has the potential to become a critical component of cloud systems due to its minimal costs and management overhead. It has been used in recent research to establish security and confidence in cloud-based applications. There are several challenges associated with the integration of blockchain technology and cloud computing:

- Blockchain was founded on the principle of decentralization in contrast to the cloud, which is entirely managed centrally and provides minimal transparency and trust configurations. Since various legal and governmental reasons make it impossible to

eliminate centralization, it is necessary to adopt a hybrid strategy whereby the cloud provider maintains some level of control while maintaining trust and transparency with the cloud users.

- Cloud data is protected from unauthorized access, while blockchain data is freely accessible. Cloud services based on blockchain technology will be widely adopted due to privacy concerns.
- Unlike cloud systems, blockchain systems are inherently susceptible to scaling problems.

The integration of blockchain technology with cloud-based services is urgently needed despite the mentioned issues. However, the question remains of how to integrate blockchain technology with the cloud. A cloud computing business model is attractive because of outsourcing services, but users and outsourcing service providers distrust one another. Zhang, et al. [27] developed the BCPay framework for fair payment for

outsourcing services in cloud computing, which achieves soundness and robustness. The system is also highly efficient regarding transaction volume and computational costs. Velmurugadass, et al. [28] developed a cloud-based Software Defined Network (SDN) for monitoring data and evidence-related operations. The SDN controller employs a blockchain system in order to protect the evidence gathered through data and user signatures. Using the Logical Graph of Evidence (LGoE), the investigator generates a report, evaluates the evidence, retrieves the evidence, and identifies the evidence. Using the evidence provided by the controller, the investigator can construct a Logical Graph of Evidence (LGoE).

Wilczyński and Kołodziej [29] proposed a blockchain-based cloud scheduler model. This model has improved the effectiveness of preparing schedules, and the simulator returns a schedule with a shorter makespan than previous individual scheduling methods. Using blockchain technology, Li, et al. [30] proposed a robust, cost-aware data caching strategy that minimizes the possibility of cache data tampering. To address trust issues between consumers, sellers, and agents, Rahman, et al. [31] propose a new exchange scheme combining blockchain with SDN to address the security risks associated with cloud computing. Distributed blockchain networks securely enable data storage and transmission, allowing scalability, flexibility, and privacy. Blockchain technology also ensures the security and privacy of data, while maintaining the system's integrity.

In order to monitor the activities of users and administrators, an audit log is essential. Still, it is susceptible to manipulation if an attacker can access the system. Attackers may modify and delete log entries or even create false entries to cover their tracks. Managing audit logs requires the protection of these records from unauthorized access. Keeping the log in a secure location with restricted access and recording and monitoring all access is essential to prevent unauthorized access.

Furthermore, the log should be regularly backed up and stored in a secure offsite location in order to detect and correct any modifications or deletions. Using blockchain technology, Ali, et al. [32] propose a Log Management System that addresses several limitations. Functionality and performance were superior to those of previous models. Xu, et al. [33] have introduced a blockchain-based method for managing cloudlets in a multi-media workflow. A multi-media application is enhanced using NSGA III, and an optimal scheduling decision is made using ELECTRE. Eltayieb, et al. [34] developed a cloud-based data-sharing platform based on blockchain technology coupled with attribute-based sign encryption. The proposed scheme complies with the security requirements of cloud computing, including confidentiality and unforgeability. The smart contract also eliminates the problem with traditional cloud servers, such as returning incorrect results.

Awadallah and Samsudin [35] introduced a cloud relational database with an enhanced structure based on blockchain technology. The client has the ability to detect and prevent errors in cloud relational database manipulation by employing a self-verification mechanism. They proposed two systems for improving the mechanism's performance: an agile blockchain-based cloud relational database and a secure blockchain-based

cloud relational database. Byzantine fault tolerance distributes both systems across several cloud service providers. The SHA-256 algorithm is also used in both systems to link records. In addition, blockchain-based cloud relational databases that operate on a proof-of-work consensus prevent data offensive operations. A blockchain-based cloud relational database is highly recommended for high throughput databases based on performance and security analysis. Cloud relational databases based on blockchain technology are recommended for containing sensitive data and performing poorly in terms of throughput. The flexibility of cloud-based relational databases allows users to operate them according to their specific requirements.

Intrusion Detection System (IDS) has become widely recognized as a valuable tool for protecting networks and information. IDS monitors network traffic and alerts administrators when malicious activities or suspicious behavior is detected. It is used to detect and prevent any unauthorized access or malicious attacks. Host IDS (HIDS) detects unauthorized use and abnormal and malicious activity on a host, whereas Network IDS (NIDS) detects network attacks and intrusions. Kumar and Singh [36] propose the development of Distributed IDS (DIDS) based on emerging and promising technologies like blockchain on a stable platform such as cloud infrastructure.

Access control is of paramount importance in cloud computing, as it is where enterprises and individuals store their sensitive data. However, the centralized access control mechanism used in the cloud poses a significant security risk. Sensitive data stored in the cloud becomes vulnerable to tampering or unauthorized disclosure by hackers or even cloud managers. This highlights the need for robust access control technologies to ensure the confidentiality and integrity of data in the cloud environment. To address this issue, Yang, et al. [37] propose AuthPrivacyChain, a blockchain-based access control framework with privacy protection. The first step is to use the account address of a node in the blockchain as the identity and simultaneously redefine the permissions for access control to the cloud. They then design processes for controlling access, authorizing users, and revoking authorizations. Lastly, the researchers implement AuthPrivacyChain using the enterprise operation system (EOS) and evaluate its performance. The results demonstrate that AuthPrivacyChain offers robust protection against unauthorized access by hackers and administrators. Additionally, it ensures the preservation of authorized privacy, providing a comprehensive security solution.

IV. DISCUSSION

In this section, we present and analyze the findings of our research on blockchain-enabled cloud security solutions. Our investigation, encompassing a systematic literature review and analysis, revealed several key insights that contribute to the evolving understanding of this field.

Cloud service users have high expectations from cloud service providers in terms of transparency, efficiency, security, and authentication of transactions, services, and applications. Trust and credibility in cloud network transactions depend on involving a trusted third party to verify, validate, and endorse

them. Incorporating business logic into the database (Ledger) and executing it becomes necessary for transaction validation and storage. Blockchain technology, with its inherent capabilities and potential, holds the promise of addressing many of the challenges faced by virtualized cloud infrastructures today. It offers secure, transparent, trustworthy, and efficient solutions for managing and registering the authorized identities of all stakeholders in the cloud. Its decentralized and distributed nature, coupled with a reliable management and governance system, allows for tracking, tracing, and effective management of cloud-related transactions. Additionally, blockchain can be leveraged to manage and store identities and services, ensuring their complete concealment from end users.

Cloud infrastructures can be made more secure by integrating blockchain technology. Oracle Blockchain Cloud Service is one example of a blockchain-enabled cloud solution currently being implemented. A virtualized cloud environment can also benefit from the use of blockchain technology. All connected cloud devices and services can be registered and identified on the blockchain ledger through a set of attributes and complex relationships. Consequently, virtualized cloud supply chain networks can provide provenance at all levels. Cloud-enabled supply chains involve numerous stakeholders, ranging from cloud infrastructure facilities, vendors, suppliers, and service providers to distributors, shippers, installers, owners, repairers, and re-installers. Anonymity is a key aspect in large-scale cloud environments. To ensure privacy and prevent third-party service providers from accessing private information, an electronic wallet is created and installed within cloud systems. Furthermore, blockchain-enabled smart contracts play a crucial role in managing, controlling, and securing cloud services and devices. The previous section highlighted the significant features of blockchain technology that are particularly valuable for cloud platforms, especially in terms of enhancing cloud security.

A virtualized cloud system offers anonymity for user information and service data, which can be strengthened by integrating blockchain-enabled solutions. One such solution is the implementation of electronic wallets within large-scale cloud environments, which utilize blockchain platforms to store users' and services' data securely. By leveraging a blockchain network, cloud service records can be stored, and the identities of cloud users and service providers can be authenticated and validated. This combination of virtualized cloud systems and blockchain technology provides an additional layer of security and trust in cloud-based transactions and interactions. Whenever a cloud service provider connects to a blockchain network, it will prove and sign its transactions cryptographically, which can be tracked and tracked by users or cloud service providers participating in the network. A blockchain-based smart contract ensures user and service privacy by controlling who has permission to update, upgrade, patch, provide new key pairs, initiate a service or repair request, and change ownership. The blockchain network provides fault tolerance and resilience to cloud users, as the failure of a single node will not affect the entire virtualized cloud infrastructure. By integrating blockchain into a cloud infrastructure solution, blockchain-as-a-service (BaaS)

can be implemented. BaaS allows users to leverage the advantages of blockchain technology without having to build an entire infrastructure from scratch. It also provides a secure platform for cloud services, and users can manage, monitor, and control their data easily. Additionally, BaaS can enable cost and time savings for businesses. Furthermore, BaaS enables businesses to scale quickly and efficiently while providing a secure and reliable platform.

Several key findings have emerged from our comprehensive exploration of blockchain-enabled cloud security approaches. Through a systematic review of the literature, we categorized and analyzed various mechanisms that leverage blockchain technology to enhance the security of cloud computing environments. Our investigation revealed that these mechanisms encompass data integrity verification, access control, consensus protocols, and auditability. These findings underscore the potential of blockchain technology to offer innovative solutions for addressing the unique security challenges inherent in cloud environments. Comparing our findings with those of previous studies, our research aligns with the insights presented by AlMuraytib, et al. [38]. Both studies highlight the potential of blockchain to enhance cloud security by ensuring data integrity and enhancing trust in cloud transactions. While we echo these sentiments, our study goes further by delving deeper into the distinct security challenges that cloud computing faces and presenting novel mechanisms specifically designed to tackle these issues. Our focus on tailoring blockchain solutions to address the intricacies of cloud security sets our work apart and contributes a fresh perspective to the field. Furthermore, our research contributes a domain-specific understanding of how blockchain principles can be adapted to meet the security demands of cloud computing. While existing studies primarily emphasize the application of general blockchain principles, our work recognizes the need for nuanced approaches to accommodate the intricacies of cloud environments. By doing so, we bridge a critical gap in the literature and offer a more targeted and effective path toward securing cloud-based systems.

V. CONCLUSION

This paper discussed the potential for transformation that arises from integrating blockchain technology with cloud computing, focusing on overcoming major security and performance challenges. The study highlighted the diverse benefits that this combination brings to the field of information technology. The initial phase of our investigation was identifying security concerns inherent in cloud computing. The need to protect data integrity arises from the vulnerability to data modification and compromise, hence requiring the development of novel solutions. Blockchain technology, known for its robust resistance to tampering, is a valuable strategic partner in this context. The originality of our study is in the application of blockchain technology specifically tailored to address security concerns in cloud computing. In contrast to prior research that use broad concepts, our research emphasizes customizing blockchain solutions to address the complexities of cloud computing, showcasing novel processes. The procedures above encompass data integrity verification, access control, and consensus protocols, effectively target and mitigate security vulnerabilities. The research findings have

far-reaching ramifications for both the academic community and several industries. By providing practitioners and researchers with security solutions tailored to certain domains, we improve the reliability of transactions conducted in cloud-based environments. Incorporating blockchain technology facilitates improving data quality and dependability, which is a crucial need in several application fields. As we contemplate the future, the convergence of cloud computing and blockchain technology presents significant possibilities. The secure and efficient data storage and processing offer potential competitive benefits for firms. The trajectory forward necessitates more investigation in order to fully elucidate the comprehensive range of repercussions stemming from this fusion. This discovery signifies the advent of a hopeful era in which cloud computing is strengthened by the powerful security framework of blockchain, leading to a trajectory towards a digital landscape that is more safe and efficient.

REFERENCES

- [1] B. Pourghbleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, pp. 1-24, 2021.
- [2] F. Nzanywayingoma and Y. Yang, "Efficient resource management techniques in cloud computing environment: a review and discussion," *International Journal of Computers and Applications*, vol. 41, no. 3, pp. 165-182, 2019.
- [3] D. C. Wyld, *Moving to the cloud: An introduction to cloud computing in government*. IBM Center for the Business of Government, 2009.
- [4] B. Pourghbleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23-34, 2017.
- [5] L. Jayashree, G. Selvakumar, L. Jayashree, and G. Selvakumar, "Cloud Solutions for IoT," *Getting Started with Enterprise Internet of Things: Design Approaches and Software Architecture Models*, pp. 31-48, 2020.
- [6] M. M. Salim, S. K. Singh, and J. H. Park, "Securing Smart Cities using LSTM algorithm and lightweight containers against botnet attacks," *Applied Soft Computing*, vol. 113, p. 107859, 2021.
- [7] V. Hayyolalam, B. Pourghbleh, and A. A. Pourhaji Kazem, "Trust management of services (TMOs): Investigating the current mechanisms," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 10, p. e4063, 2020.
- [8] V. Hayyolalam, B. Pourghbleh, A. A. P. Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 471-498, 2019.
- [9] M. E. Hussain and R. Hussain, "Cloud Security as a Service Using Data Loss Prevention: Challenges and Solution," in *Internet of Things and Connected Technologies: Conference Proceedings on 6th International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, 2021, 2022: Springer, pp. 98-106.
- [10] A. Hedhli and H. Mezni, "A survey of service placement in cloud environments," *Journal of Grid Computing*, vol. 19, no. 3, pp. 1-32, 2021.
- [11] D. Hazra, A. Roy, S. Midya, and K. Majumder, "Energy aware task scheduling algorithms in cloud environment: A survey," in *Smart Computing and Informatics*: Springer, 2018, pp. 631-639.
- [12] J. Dizdarević, F. Carpio, A. Jukan, and X. Masip-Bruin, "A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration," *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1-29, 2019.
- [13] E. J. Ghomi, A. M. Rahmani, and N. N. Qader, "Load-balancing algorithms in cloud computing: A survey," *Journal of Network and Computer Applications*, vol. 88, pp. 50-71, 2017.
- [14] L. M. Dang, M. Piran, D. Han, K. Min, and H. Moon, "A survey on internet of things and cloud computing for healthcare," *Electronics*, vol. 8, no. 7, p. 768, 2019.
- [15] T. Gera, J. Singh, A. Mehbodniya, J. L. Webber, M. Shabaz, and D. Thakur, "Dominant feature selection and machine learning-based hybrid approach to analyze android ransomware," *Security and Communication Networks*, vol. 2021, pp. 1-22, 2021.
- [16] S. N. H. Bukhari, J. Webber, and A. Mehbodniya, "Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates," *Scientific Reports*, vol. 12, no. 1, p. 7810, 2022.
- [17] J. Webber, A. Mehbodniya, Y. Hou, K. Yano, and T. Kumagai, "Study on idle slot availability prediction for WLAN using a probabilistic neural network," in *2017 23rd Asia-Pacific Conference on Communications (APCC)*, 2017: IEEE, pp. 1-6.
- [18] C. B. Tan, M. H. A. Hijazi, Y. Lim, and A. Gani, "A survey on proof of retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends," *Journal of Network and Computer Applications*, vol. 110, pp. 75-86, 2018.
- [19] P. Yang, N. Xiong, and J. Ren, "Data security and privacy protection for cloud storage: A survey," *IEEE Access*, vol. 8, pp. 131723-131740, 2020.
- [20] M. Rady, T. Abdelkader, and R. Ismail, "Integrity and confidentiality in cloud outsourced data," *Ain Shams Engineering Journal*, vol. 10, no. 2, pp. 275-285, 2019.
- [21] T. Wang, Q. Yang, X. Shen, T. R. Gadekallu, W. Wang, and K. Dev, "A privacy-enhanced retrieval technology for the cloud-assisted internet of things," *IEEE transactions on industrial informatics*, vol. 18, no. 7, pp. 4981-4989, 2021.
- [22] V. Indić, M. Kovačević, M. Simić, and G. Sladić, "Towards Local Cloud Infrastructure in Developing Countries as a Response to Data Localization Regulations," ed: ICIST, 2022.
- [23] S. Mohapatra, N. Bajpai, T. Swarnkar, and M. Mishra, "Raw Data Redundancy Elimination on Cloud Database," in *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2020*, 2020: Springer, pp. 395-405.
- [24] J. Doyle, M. Golec, and S. S. Gill, "Blockchainbus: A lightweight framework for secure virtual machine migration in cloud federations using blockchain," *Security and Privacy*, vol. 5, no. 2, p. e197, 2022.
- [25] A. Alkhateeb, C. Catal, G. Kar, and A. Mishra, "Hybrid blockchain platforms for the internet of things (IoT): A systematic literature review," *Sensors*, vol. 22, no. 4, p. 1304, 2022.
- [26] I. Yaqoob, K. Salah, R. Jayaraman, and Y. Al-Hammadi, "Blockchain for healthcare data management: opportunities, challenges, and future recommendations," *Neural Computing and Applications*, pp. 1-16, 2021.
- [27] Y. Zhang, R. H. Deng, X. Liu, and D. Zheng, "Blockchain based efficient and robust fair payment for outsourcing services in cloud computing," *Information Sciences*, vol. 462, pp. 262-277, 2018.
- [28] P. Velmurugadass, S. Dhanasekaran, S. S. Anand, and V. Vasudevan, "Enhancing Blockchain security in cloud computing with IoT environment using ECIES and cryptography hash algorithm," *Materials Today: Proceedings*, vol. 37, pp. 2653-2659, 2021.
- [29] A. Wilczyński and J. Kołodziej, "Modelling and simulation of security-aware task scheduling in cloud computing based on Blockchain technology," *Simulation Modelling Practice and Theory*, vol. 99, p. 102038, 2020.
- [30] C. Li, S. Liang, J. Zhang, Q.-e. Wang, and Y. Luo, "Blockchain-based data trading in edge-cloud computing environment," *Information Processing & Management*, vol. 59, no. 1, p. 102786, 2022.
- [31] A. Rahman, M. J. Islam, S. S. Band, G. Muhammad, K. Hasan, and P. Tiwari, "Towards a blockchain-SDN-based secure architecture for cloud computing in smart industrial IoT," *Digital Communications and Networks*, 2022.
- [32] A. Ali, A. Khan, M. Ahmed, and G. Jeon, "BCALS: Blockchain-based secure log management system for cloud computing," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 4, p. e4272, 2022.
- [33] X. Xu, Y. Chen, Y. Yuan, T. Huang, X. Zhang, and L. Qi, "Blockchain-based cloudlet management for multi-media workflow in mobile cloud

- computing," *Multi-media Tools and Applications*, vol. 79, pp. 9819-9844, 2020.
- [34] N. Eltayieb, R. Elhabob, A. Hassan, and F. Li, "A blockchain-based attribute-based signcryption scheme to secure data sharing in the cloud," *Journal of Systems Architecture*, vol. 102, p. 101653, 2020.
- [35] R. Awadallah and A. Samsudin, "Using blockchain in cloud computing to enhance relational database security," *IEEE Access*, vol. 9, pp. 137353-137366, 2021.
- [36] M. Kumar and A. K. Singh, "Distributed intrusion detection system using blockchain and cloud computing infrastructure," in *2020 4th international conference on trends in electronics and informatics (ICOEI)(48184)*, 2020: IEEE, pp. 248-252.
- [37] C. Yang, L. Tan, N. Shi, B. Xu, Y. Cao, and K. Yu, "AuthPrivacyChain: A blockchain-based access control framework with privacy protection in cloud," *IEEE Access*, vol. 8, pp. 70604-70615, 2020.
- [38] S. AlMuraytib, L. Alqurashi, and S. Snoussi, "Blockchain-based solutions for Cloud Computing Security: A Survey," in *Proceedings of the 6th International Conference on Future Networks & Distributed Systems*, 2022, pp. 338-342.

Research on Enterprise Supply Chain Anti-Disturbance Management Based on Improved Particle Swarm Optimization Algorithm

Tongqing Dai

Business School, Huanggang Normal University, Huanggang, 438000, China

Abstract—A supply chain that is effective and of the highest caliber boosts customer happiness as well as sales and earnings, increasing the company's competitiveness in the market. It has been discovered that the standard supply chain management technique leaves the supply chain with weak supply chain stability because it has a low ability to withstand the manufacturer's production behaviour. An enterprise supply chain resistance management model is built using the study's proposed particle swarm optimisation technique, which is based on a genetic algorithm with stochastic neighbourhood structure, to solve this issue. The suggested technique outperformed the other two algorithms utilised for comparison in a performance comparison test, with a stable particle swarm fitness value of 0.016 after 800 iterative iterations and the fastest convergence. The proposed model was then empirically examined, and the results revealed that the production team using the model completed the same volume of orders in 32 days while making \$460,000 more in profit. With scores of 4.5, 4.5, 4.3, 4.3, 4.2, and 4.2, respectively, the team also had the lowest values of the six forms of employee anti-production conduct, outperforming the comparative management style. In summary, the study proposes an anti-disturbance management model for enterprise supply chains that can rationalise the scheduling of manufacturers' production behaviour and thus improve the stability of the supply chain.

Keywords—Supply chain; particle swarm optimization algorithm; genetic algorithm; inverse production behaviour; neighbourhood structure

I. INTRODUCTION

With the development of the global economy and the intensification of competition, the supply chain of enterprises is faced with more and more disturbances and uncertainties [1]. These disturbances include fluctuations in market demand, changes in raw material prices, natural disasters, etc. which have a huge impact on the operation of the supply chain [2-3]. Therefore, enterprises need to strengthen the anti-disturbance ability of the supply chain to ensure the efficient operation and stability of the supply chain. In supply chain anti-disturbance management, it is an important task to optimize resource allocation and decision-making in the supply chain [4-5]. The traditional particle swarm optimization algorithm often ignores the nonlinear relations and complex constraints in the supply chain, resulting in unstable optimization results and difficult to be applied in practice. Therefore, a new optimization algorithm is needed to solve this problem. In order to solve this problem and improve the anti-disturbance

ability of enterprise supply chain, this paper proposes to improve the particle swarm optimization algorithm by using genetic algorithm and random neighborhood structure, and build the anti-disturbance management model of enterprise supply chain based on the improved algorithm. It is hoped that this model can improve the anti-interference ability of supply chain, improve the stability of supply materials, and enhance the market competitiveness of enterprises. The research is of great significance for enterprises to improve supply chain anti-disturbance ability, improve operation efficiency and reduce operation cost. At the same time, the anti-disturbance management method of enterprise supply chain based on improved algorithm also has certain theoretical and practical value, and has certain reference significance for the application and promotion of optimization algorithm. The innovation point of the research is that, considering various variables and uncertainties of the supply chain, genetic algorithm and random neighborhood structure are used to improve the traditional PSO based optimization algorithm, so as to optimize the resource allocation and decision-making in the supply chain, so as to improve the robustness and flexibility of the supply chain. The second section of this research is the in-depth study of the application of particle swarm optimization algorithm and supply chain problems in recent years. The third section analyzes the problems existing in the classical particle swarm optimization algorithm, proposes to improve it by using genetic algorithm and random neighborhood structure, and establishes the enterprise supply chain anti-disturbance management model based on the improved algorithm. The fourth section is the performance comparison test of the improved algorithm proposed in the research, and the practical application effect analysis of the enterprise supply chain anti-disturbance management model. The fifth section is the summary and conclusion of the whole research.

II. RELATED WORKS

A particle swarm optimisation technique for multi-objective solutions has been researched by scientists as part of the ongoing advancement of science and technology, and it is now widely employed in many neighbourhoods. To solve the issues with the design of steel pipe support weighing structures, Zakian et al. suggested an optimisation algorithm merging particle swarm algorithm with grey wolf optimiser. The findings of the empirical investigation suggest that the optimisation technique can enhance the structural. The outcomes demonstrated that the optimisation approach might

enhance the pipe support structure's structural load-bearing and dimensional binding performance [6]. In an effort to address the issue that thermal coupling can lower the control accuracy of eccentric rotor extruders, Wen et al. proposed a control algorithm based on the particle swarm optimisation algorithm and neuron proportional integral differentiation [7]. After comparison tests, it was found that the algorithm can offset the effect of thermal coupling and improve the control accuracy of eccentric rotor extruders. To solve the issue that it is challenging to locate global peaks of PV arrays under shading conditions, Javed's team suggested a particle swarm optimisation approach in conjunction with adaptive learning. The results of comparative experimental study indicate that the algorithm outperforms conventional algorithms in terms of convergence speed and success rate, operating with an average efficiency of 99.65% [8]. To address the issue of low vehicle guidance accuracy in congested urban networks, Zouari's team suggested a hierarchical interval type 2 fuzzy logic model based on particle swarm optimisation. After conducting a simulation test and analysing the findings, Liu et al. suggested a node localization approach based on the combination of the particle swarm optimisation algorithm and the monkey algorithm. The results indicated that the method produced improved localization effects in terms of node rate and node density [10].

A solution using human capital and digital management was put out by Song et al. to address the issue of high retailer volatility in supply chain integration. According to the empirical analysis's findings, merchants who used this model had a better long-term investment mindset, which improved the supply chain's stability [11]. To solve the problem that green sensitivity has a significant impact on the stability of green supply chains, Long's team created an evolutionary game model that incorporates the green sensitivity of the government, businesses, and consumers. Following empirical analysis, the findings demonstrated that the model can combine the three elements to create a reasonable green sensitivity, upholding the stability of the green supply chain [12]. The outcomes of the comparison experiments demonstrated that this method can increase both the speed and security of hydrogen storage [13]. To attempt to address the issue of supply-demand mismatch in the supply chain for influenza vaccines, Lin et al. proposed a segmented linear function-based procurement model. After comparative experimental analysis, it was demonstrated that the model could coordinate the supply chain for influenza vaccines more effectively by fully taking into account the supply-demand relationship in the supply chain [14]. The results of a comparative experimental examination of a multi-stage mixed integer planning model revealed that it may not only lower logistics costs in the forest supply chain but also satisfy the wood demand of the sector [15].

In conclusion, merging supply chain research with the superiority of particle swarm optimisation algorithms has been shown in a number of localities, and it is thought that doing so has some research worth. Few academics have merged the two, thus this work uses the particle swarm optimisation algorithm to the creation of an anti-disturbance management model for company supply chains in an effort to close the gap in this

research direction. It is believed that this research would increase the ability of business supply networks to withstand disruptions, lay the groundwork for increased business competitiveness, and serve the field of business supply chain management with research data.

III. RESEARCH ON ANTI-INTERFERENCE MANAGEMENT MODEL OF ENTERPRISE SUPPLY CHAIN BASED ON IMPROVED PARTICLE SWARM OPTIMIZATION ALGORITHM

It is impossible to emphasise how important the supply chain is to business operations, however traditional supply chain management is ineffective at controlling manufacturers' production behaviours, causing the supply chain to be impacted by this issue and having poor stability. This chapter will build an anti-disruption management model for enterprise supply chains based on the improved algorithm by using genetic algorithm and stochastic neighbourhood structure to address the shortcomings of the conventional particle swarm optimisation algorithm.

A. Improved Particle Swarm Optimization Algorithm based on Genetic Algorithm and Random Neighborhood Structure

Particle Swarm Optimization (PSO) is a population-based stochastic optimization technique that has strong adaptive and global convergence capabilities [16]. A particle in the PSO algorithm represents a process ordering in a production plant, i.e. a feasible solution to the production scheduling scheme. The particles are initialised and the equations are shown in Eq. (1) and (2).

$$v_{i,j}(t+1) = \omega v_{i,j}(t) + c_1 r_1 [p_{i,j} - x_{i,j}(t)] + c_2 r_2 [p_{g,j} - x_{i,j}(t)] \quad (1)$$

In Eq. (1), $v_{i,j}(t+1)$ denotes the velocity of the particle at the moment $t+1$ and t denotes the moment t . $x_{i,j}$ denotes the vector of real numbers and $v_{i,j}$ denotes the velocity vector. $p_{i,j}$ and $p_{g,j}$ both denote the current optimal position of the particle. ω denotes the habituation factor and c_1 and c_2 both denote the learning factor. r_1 and r_2 denote the random number between production.

$$x_{i,j}(t+1) = x_{i,j}(t) + v_{i,j}(t+1), j \in \{1, 2, 3, \dots, n\} \quad (2)$$

In Eq. (2), $x_{i,j}(t+1)$ denotes the particle position iteration at this moment of $t+1$. In the PSO algorithm, information sharing is constructed based on the best position of individual particles with the best position information of the particle population, passing the information to other particles in the search space. Due to the lack of information exchange between each particle, the optimal position of the particle itself changes as the search for the optimal position is continuously updated, and thus the optimal position of the particle population changes as well. Genetic algorithms are a method of searching for optimal solutions that mimics the natural evolutionary process, using processes such as chromosomal gene mutation and crossover [17]. The algorithm can optimise complex problems faster than

traditional optimisation algorithms, and this study will use the mutation and crossover processes of the algorithm to change the problem of untimely information interaction in the PSO algorithm to ensure that each particle in the PSO algorithm can interact with each other to achieve information sharing and prevent local search [18]. The mutation operation in the algorithm is shown in Fig. 1.

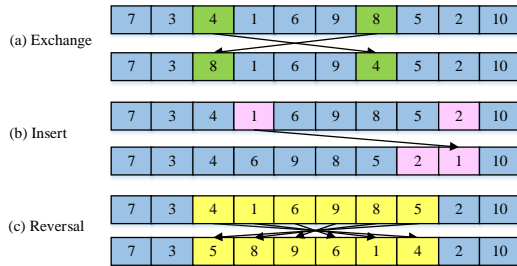


Fig. 1. Schematic representation of the variant operation.

As shown in Fig. 1, the mutation operation is performed separately for the historical best position of the particle and the current best position of the particle population. Different particles have different mutation probabilities, but the mutation operator is selected with the same probability, i.e. the selection of swap, insertion and inversion is the same. Swap variation involves swapping the positions of two randomly selected positions in the particle vector. The insertion variant is a random selection of two positions in a particle vector, which are compared in terms of their numerical size, with the smaller numerical particle being inserted after the larger value. The reversal variation involves randomly selecting two positions in the particle vector and reordering the other particles between the two positions in reverse, i.e. [4,1,6,9,8,5] becomes [5,8,9,6,1,4], as in Fig. 1(c). The crossover operation in the algorithm is shown in Fig. 2.

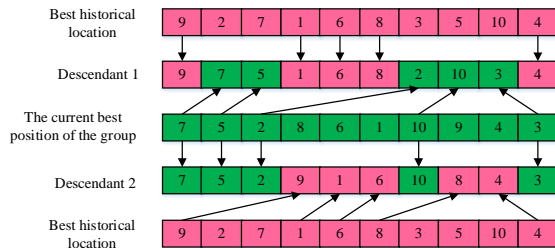


Fig. 2. Schematic diagram of the crossover operation.

As shown in Fig. 2, the historical best positions of the particles after mutation and the current best positions of the particle population are labelled and each is used as a parent to perform the crossover operator. A number of $n'(n' < n)$ positions are randomly selected from the particle historical best positions to form array 1, which is copied directly in the first child according to its position in the parent. The first offspring is directly copied in the first offspring according to its position in the parent, thus completing the first offspring. The second offspring is formed similarly. By means of mutation and crossover operations, it is possible to interact with the information between individual particles in the PSO algorithm and to improve the search accuracy of the PSO algorithm. Although the PSO algorithm is fast in global convergence, local convergence faces the problem of falling into local traps.

To solve this problem, a random neighbourhood structure containing insertion neighbourhood, exchange neighbourhood and block exchange neighbourhood was studied [19]. The insertion operation of this neighbourhood structure is highly random, making the neighbourhood structure flexible and guiding the particles to search quickly during production scheduling. The swap operation is decentralised and can lead particles to jump out of local search. Therefore, using this neighbourhood to improve the PSO algorithm not only solves the problem of the algorithm falling into local traps, but also improves the speed and accuracy of the search. The search mechanism of the random structured neighbourhood is to randomly perform one or more insert neighbourhood, swap neighbourhood and block swap neighbourhood operations, the mathematical structure of which is expressed in Eq. (3).

$$f_N = \left(M \oplus \left(C_{pm} \otimes X_{best(g)} \right) \right) \quad (3)$$

In Eq. (3), M denotes the step size for performing the neighbourhood operation; $X_{best(g)}$ denotes the current best position of the X population; and C_{pm} denotes the probability of the diameter neighbourhood operation, which is calculated as shown in Eq. (4).

$$C_{pm} = \begin{cases} (\alpha_1 \leq rand() \leq \beta_1) \Rightarrow insert(\pi, k_1, k_2) \\ (\alpha_2 \leq rand() \leq \beta_2) \Rightarrow swap(\pi, k_1, k_2) \\ (\alpha_3 \leq rand() \leq \beta_3) \Rightarrow blockswap(\pi, B_1, B_2) \end{cases} \quad (4)$$

In Eq. (4), $[\alpha_1, \beta_1]$, $[\alpha_2, \beta_2]$ and $[\alpha_3, \beta_3]$ denote three probability intervals respectively; $rand()$ denotes a random distribution between (0,1); π denotes a scheduling scheme; k denotes an artifact; B denotes a set consisting of two adjacent artifacts; $insert(\pi, k_1, k_2)$ denotes an insertion neighbourhood; $swap(\pi, k_1, k_2)$ denotes an exchange neighbourhood; and $blockswap(\pi, B_1, B_2)$ denotes a block exchange neighbourhood. The parts of the probability region that overlap each other are defined as $COM < I, S, BS >$, and when $rand()$ is between $COM < I, S, BS >$, then the operation of the random neighbourhood is pressed as $insert(\pi, k_1, k_2)$, $swap(\pi, k_1, k_2)$, $blockswap(\pi, B_1, B_2)$. The working schematic of this neighbourhood structure is shown in Fig. 3.

As shown in Fig. 3(a), the position of workpiece 1 in the scheduling plan is randomly determined in $insert(\pi, k_1, k_2)$ and randomly inserted after the position of workpiece 2. As shown in Fig. 3(b), the positions of workpiece 1 and workpiece 2 in the scheduling plan are randomly swapped in $swap(\pi, k_1, k_2)$. The positions of workpiece set 1 and workpiece set 2 are randomly swapped in $blockswap(\pi, B_1, B_2)$ as shown in Fig. 3(c). Combining the above, the PSO algorithm is improved using genetic algorithm with random neighbourhood structure, and the improved algorithm is defined as HPSO-R.

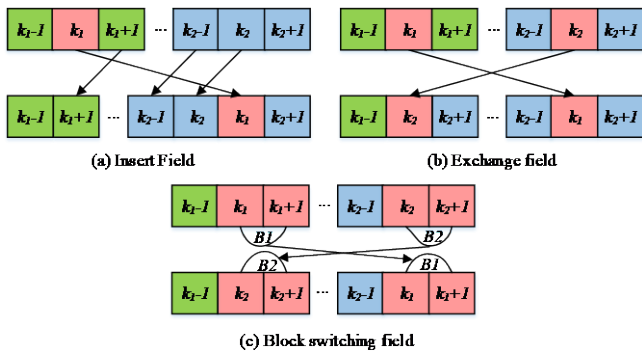


Fig. 3. A Schematic diagram of the domain structure.

B. Construction of Supply Chain Anti-Disturbance Management Model based on Improved Particle Swarm Optimization Algorithm

The supply chain is a network chain structure formed by the whole process of sending products from production to consumers. The supply chain of an enterprise is the lifeblood of the enterprise economy and its stable operation is very important for the development of the enterprise. The schematic diagram of the supply chain structure is shown in Fig. 4.

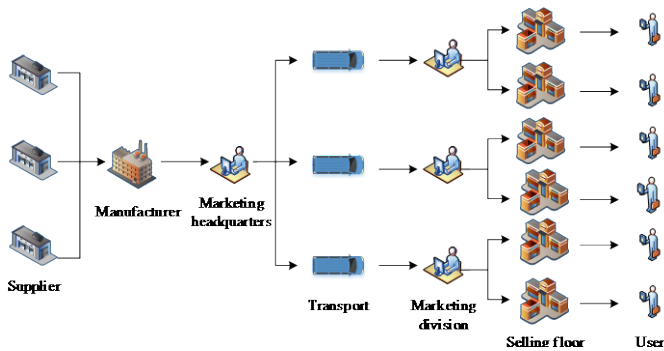


Fig. 4. Supply Chain Schematic diagram.

The stability of the enterprise supply chain is mainly influenced by the stability of the products produced by the manufacturer. The study will optimise the manufacturer's production management scheduling model through the HPSO-R algorithm to construct an anti-disturbance management model for the enterprise supply chain. The reasonableness of manufacturers' production management scheduling can be reflected by the counterproductive behaviours generated by employees [20]. Anti-production behaviours are divided into six categories, the first of which is job satisfaction, the equation for which is shown in Eq. (5).

$$L_1 = INTEG(R_1 - R_2, 2) \quad (5)$$

In Eq. (5), the letters L_1 , R_1 , and R_2 stand for work satisfaction, rate of increase in satisfaction, and rate of drop in satisfaction, respectively. The equation for the second category, organisational justice, is given in Eq. (6).

$$L_2 = INTEG(R_3 - R_4, 1) \quad (6)$$

In Eq. (6), the letters L_2 , R_3 , and R_4 stand for the

organisational sense of justice, the rate at which it is increasing, and the rate at which it is decreasing. Eq. (7)'s equation for the third category, "sense of team climate," is shown.

$$L_3 = INTEG(R_5 - R_6, 1.5) \quad (7)$$

In Eq. (7), L_3 stands for the team's overall environment, R_5 for the rate at which it is improving, and signifies the rate at which it is deteriorating. The level of supervision, which is the fourth category, is determined as illustrated in Eq. (8).

$$L_4 = L_2 \times A_1 + L_6 \times A_2 + L_7 \times A_3 + L_8 \times A_4 \quad (8)$$

L_4 represents the degree of supervision, L_6 represents the degree of organisational culture building, L_7 represents the number of behavioural corrections, L_8 represents the perfection of the supervision mechanism, and $A_i (i = 1, 2, \dots, n)$ represents the weighting factor in Eq. (8). Level of group regulation, the fifth category, has an equation that is represented in Eq. (9).

$$L_5 = INTEG(R_7 - R_8, 1.5) \quad (9)$$

In Eq. (9), L_5 stands for the group normative level, R_7 for the rate of increase of the group normative level, and R_8 for the rate of fall of the group normative level. The level of organisational culture building, the equation for which is stated in Eq. (10), is the sixth category.

$$L_6 = INTEG(R_9, 2) \quad (10)$$

In Eq. (10), L_6 stands for the organisational culture's level of development, and R_9 for its pace of growth. Eq. (11) displays the equation for determining anti-productive behaviours.

$$L_7 = L_1 \times F_1 + L_2 \times F_2 + \dots + L_6 \times F_6 \quad (11)$$

L_7 stands for unproductive behaviour, and $F_i (i = 1, 2, \dots, 6)$ stands for the coefficients of each of the six contributing elements described above in Eq. (11). Fig. 5 illustrates the interrelationship of the variables driving employee unproductive behaviour.

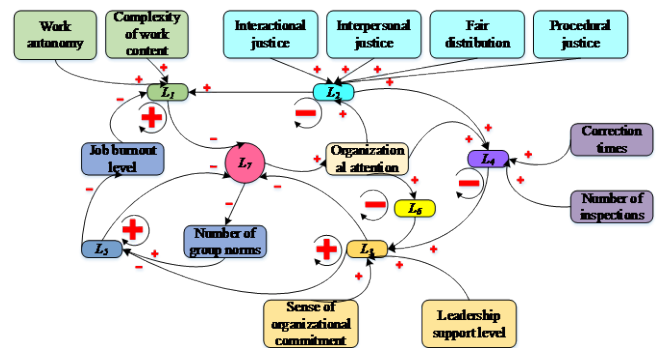


Fig. 5. Plot of anti-productive behavior factors

In the manufacturer's production management scheduling model, the initial scheduling target is calculated as shown in Eq. (12).

$$\min \left\{ f(\pi_0^s) = \sum_{j=1}^n w_j \cdot C_j, f(\pi_0^m) = \sum_{j=1}^n w'_j \cdot C'_j \right\} \quad (12)$$

In Eq. (12), j denotes the workpiece; π_0^s is the supplier's initial dispatch time; π_0^m is the manufacturer's initial dispatch time; w_j and w'_j are the supplier's weighting factor and the manufacturer's weighting factor respectively; BBB and C_j are the supplier's completion time and the manufacturer's completion time respectively. To strengthen the resilience to disturbances in the supply chain, the study introduces disturbance management theory to optimise the model. Interference management models the optimisation of individual practical problems and disturbance events, e.g. in the face of machine downtime during a manufacturer's production process, the interference management scheduling objective optimisation equation is shown in Eq. (13).

$$\left\{ \begin{aligned} \min \left\{ f_1(\pi') = \sum_{j=1}^n w'_j \cdot C'_j, f_2(\pi') = \sum_{j=1}^n w'_j \cdot \bar{\Delta}t'_0 \right\} \\ \bar{\Delta}t'_0 = \max \{ C'_j - \bar{C}'_j, 0 \} \end{aligned} \right. \quad (13)$$

In Eq. (13), $f_1(\pi')$ denotes the optimisation objective of the manufacturer's disturbance repair scheme as well as the initial scheduling scheme, $f_2(\pi')$ denotes the minimisation objective and \bar{C}'_j denotes the manufacturer's completion time of the workpiece in the initial scheduling scheme. During the scheduling arrangement of the production product, the cost benefit between the supplier and the manufacturer also needs to be considered, i.e. the objective of maximising the benefits of cooperation, which is calculated as shown in equations (14) and (15).

$$\min \{ f_3(\pi') = -V_m \cdot V_s \} \quad (14)$$

In Eq. (14), V_m denotes the manufacturer's revenue after the disturbance, and V_s denotes the supplier's revenue after the disturbance.

$$\left\{ \begin{aligned} D_j^s \leq S_j^m \\ S_j^s, C_j^s \notin [t_1, t_2], (j \in list) \\ (S_j \geq C_k) \vee (S_k \geq C_j), \forall j, k \in J \end{aligned} \right. \quad (15)$$

In Eq. (15), D_j^s denotes the supplier's delivery time; S_j^m denotes the manufacturer's processing start time; C_j^s and C_j^m denote the supplier's processing time and completion time respectively. For the above multi-objective optimisation problem of initial scheduling objective, disturbance management scheduling objective and cooperation revenue

maximisation objective, the research proposed HPSO-R algorithm can be used to find the optimal solution for it. The workflow is shown in Fig. 6.

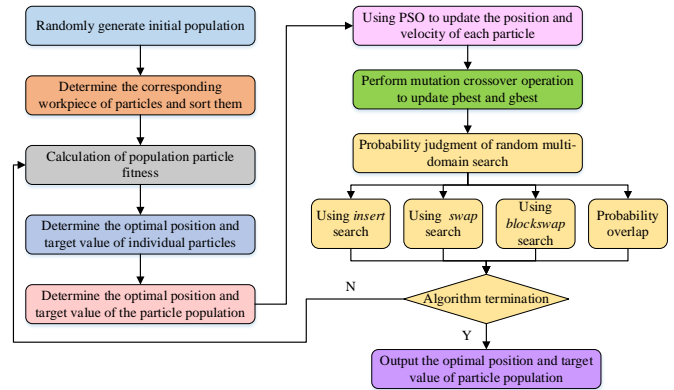


Fig. 6. Workflow of the HPSO-R algorithm.

As shown in Fig. 6, the HPSO-R algorithm is made from the basic PSO algorithm, improved by variation, crossover operations and random neighbourhood structure in the genetic algorithm, and has a higher search accuracy than the traditional PSO algorithm, with superiority in both global search as well as local search. The work flow of the algorithm is as follows: the input data is randomly generated to generate the initial population, and then the fitness value of the population is calculated according to the corresponding workpiece. According to the fitness value, the optimal position and target value of the individual particle are determined, and then the optimal position and target value of the population are determined. After that, the position and velocity of each particle are updated by particle swarm optimization algorithm, and the optimal position and target value of individual particle and the optimal position and target value of group are updated by mutation and cross operation in genetic algorithm. The mutation operation of genetic algorithm is divided into three types, namely exchange mutation, insert mutation and reverse mutation, and the best position of the particle history and the best position of the particle population are respectively changed. The crossover operation is to mark the historical best position of the particle and the current best position of the particle population after the mutation operation, and perform the crossover operator as the parent respectively, and realize the transformation of the historical best position of the particle and the current best position of the particle population through genetics. Through variation and cross operation, the information exchange between individual particles in PSO algorithm can be realized, and the search accuracy of PSO algorithm can be improved. Then the random structure neighborhood is used to perform insertion neighborhood, exchange neighborhood, block exchange neighborhood and probability overlap operations. The insertion operation of the neighborhood structure has a strong randomness, which makes the neighborhood structure change flexibly, and leads the particle to search quickly in the production scheduling process. The exchange operation is decentralized and can guide the particles out of the local search. Using this field can not only avoid PSO algorithm falling into local traps, but also improve the search speed and accuracy of the algorithm. Finally, it is

judged whether the optimal position and target value of the particle swarm after searching meet the termination condition of the algorithm, and if so, the result is output. If it is not satisfied, the fitness value of the new population is calculated according to the result, and the previous steps are repeated until the algorithm termination condition is satisfied, and the optimal position and target value of the particle swarm are output. In summary, the improved algorithm not only has high speed and high accuracy, but also will not fall into the local optimal situation, which can strengthen the anti-interference ability of enterprise supply chain.

IV. RESULTS AND DISCUSSION

To verify the performance of the HPSO-R algorithm proposed in the study, this study will conduct a comparison test using Visual software and the PSO algorithm, Genetic Algorithm Trade Off Model (GA-TOM) algorithm will be used as the comparison algorithm. The experiment uses the metrics of the Overall Nondominated Vector Generation (ONVG), the Uniformity of Distribution of Non-inferior Solutions (UDNS), the Non Inferior Solution Dominance Ratio (NISD) and the Average Distance between the Non-inferior Solution and the Optimal Pareto Front (ADF) to evaluate the algorithm in a comprehensive manner. An empirical analysis of the research's proposed improved algorithm-based model for the anti-disturbance management of an enterprise's supply chain is then carried out, with workers divided into three groups within a small factory, using the HPSO-R model, the PSO model and the GA-TOM model in a two-month comparative trial. The models will be comprehensively evaluated using indicators such as product completion time and profit in the trial.

A. Comparative Analysis of Performance of Improved Particle Swarm Optimization Algorithm

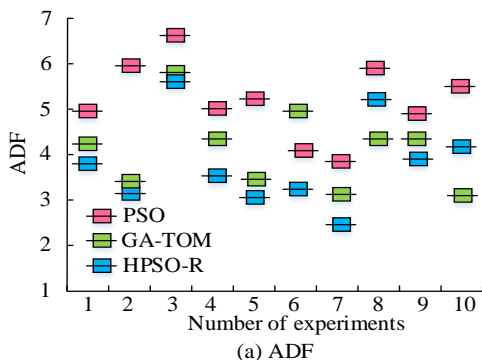


Fig. 8. A for ADF and DPNS for the three algorithms.

As shown in Fig. 8(a), the ADF value of the HPSO-R algorithm is 5.6 at maximum, 2.4 at minimum and 4.1 at average; the ADF value of the GA-TOM algorithm is 5.8 at maximum, 3.1 at minimum and 4.6 at average; and the ADF value of the PSO algorithm is 6.6 at maximum, 4.8 at minimum and 5.8 at average. The ADF value of the HPSO-R algorithm is lower than the other two. The ADF values of the HPSO-R algorithm are lower than the other two algorithms. As shown in Fig. 8(b), the NISD value of the HPSO-R algorithm was 0.62 at maximum, 0.11 at minimum, and 0.31

at mean; the NISD value of the GA-TOM algorithm was 0.35 at maximum, 0.11 at minimum, and 0.25 at mean; the NISD value of the PSO algorithm was 0.42 at maximum, 0.0 at minimum, and 0.06 at mean. The NISD value of the HPSO-R algorithm was higher than the other two algorithms. The NISD values of the HPSO-R algorithm were higher than those of the other two algorithms. In summary, the HPSO-R algorithm outperformed the comparison algorithms in terms of ADF and NISD. The experiment reflects the convergence speed of the algorithm by recording the change curve of the particle swarm

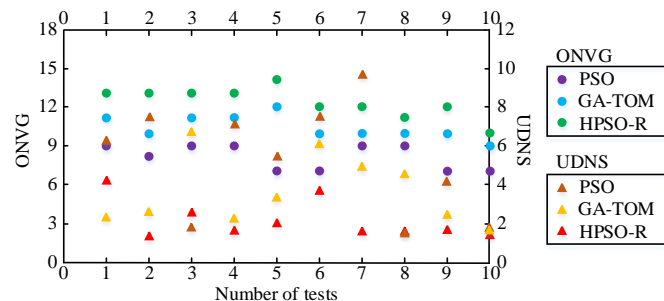
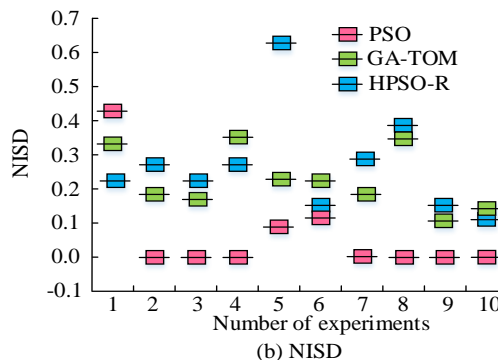


Fig. 7. ONVG and UDNS for the three algorithms.

As shown in Fig. 7, the ONVG values of the HPSO-R algorithm are 13, 13, 13, 13, 14, 12, 12, 11, 12, 10; the ONVG values of the GA-TOM algorithm are 11, 10, 11, 11, 12, 10, 10, 10, 10, 9; and the ONVG values of the PSO algorithm are 9, 8, 9, 9, 7, 7, 9, 9, 7. The minimum UDNS values of the HPSO-R, GA-TOM and PSO algorithms are 1.6, 1.8 and 1.9, respectively, and the maximum UDNS values are 4.2, 6.8 and 9.6. The overall UDNS values of the HPSO-R algorithm are smaller than those of the other two algorithms. In summary, the HPSO-R algorithm outperforms the other two compared algorithms in terms of ONVG and UDNS, two evaluation metrics. The ADF and NISD test results of the three algorithms are shown in Fig. 8, where the smaller the ADF value the better the performance, and the larger the NISD value the stronger the performance.



fitness value during the iterative operation of the algorithm. The smaller the swarm fitness value, the smaller the difference between the result and the optimal solution, and the better the performance of the algorithm. The curves of particle swarm fitness values for the three algorithms are shown in Fig. 9.

According to Fig. 9, which compares the particle swarm fitness curves of the three algorithms, the PSO algorithm begins to stabilise at 0.029 after 200 iterations, followed by the GA-TOM algorithm at 0.021 after 150 iterations, and the HPSO-R algorithm at 0.016 after 100 iterations. The HPSO-R method outperforms the comparative algorithms in terms of performance and has the fastest convergence speed and lowest particle swarm fitness value. In conclusion, the study's evaluation metrics showed that the proposed HPSO-R algorithm performed better than the other two comparative algorithms, proving its superiority.

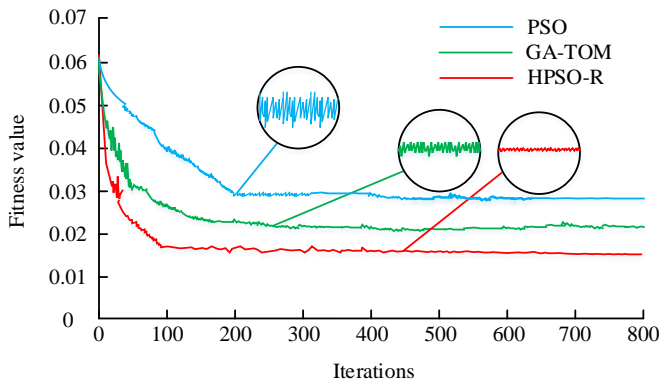


Fig. 9. The particle population fitness value curve of the three algorithms.

B. Analysis of the Effectiveness of the Application of an Enterprise Supply Chain Anti-Disturbance Management Model

In the experiments on the practical application of the anti-disturbance management model of the enterprise supply chain proposed by the study, three groups of employees using the HPSO-R model, the PSO model and the GA-TOM model were assigned the same number of orders, and the experimental results on the order completion time as well as the overall profit are shown in Fig. 10.

From Fig. 10(a), it can be seen that for the same number of orders, the completion time was 59 days for the group using

the PSO model, 43 days for the group using the GA-TOM model and 32 days for the group using the HPSO-R model. From Fig. 10(b), it can be seen that the profits of these three groups using the PSO model, the GA-TOM model and the HPSO-R model after completing the same number of orders are \$340,000, \$380,000 and \$460,000 respectively. The group using the HPSO-R model had the shortest time to complete orders and the highest profit, all better than the comparison model. As the value of employee-generated counterproductive behaviour can reflect the rationality of a manufacturer's production management scheduling, the six types of counterproductive behaviour of employees within the three groups using the model were recorded and their experimental results are shown in Fig. 11.

As it can be seen from Fig. 11, all three groups of employees reached their highest values of counterproductive behaviour at around day 20 of the trial. The highest values for the six categories of counterproductive behaviour were 7.1, 7.0, 6.8, 6.7, 6.4 and 6.1 in the group using the PSO model; 5.6, 5.5, 5.3, 5.2, 5.1 and 5.0 in the group using the GA-TOM model; and The HPSO-R model group had the lowest production behaviour values, indicating that the group had the most rational production management scheduling. A sample of completed goods from each group was checked and the pass rate of goods was counted, while manufacturers were invited to rate the perception of using the three models out of 10. The results of the experiment are shown in Fig. 12.

As it can be seen from Fig. 12(a), at one month into the trial, the commodity pass rate was 65% with a manufacturer rating of 6.3 in the PSO model group, 75% with a manufacturer rating of 7.1 in the GA-TOM model group, and 83% with a manufacturer rating of 8.2 in the HPSO-R model group. As it can be seen from Fig. 12(b), at two months into the trial, the PSO model group had a commodity qualification rate of 66% and a manufacturer rating of 6.6, the GA-TOM model group had a commodity qualification rate of 78% and a manufacturer rating of 7.9, and the HPSO-R model group had a commodity qualification rate of 92% and a manufacturer rating of 9.1. Combining the above experimental results, it can be seen that the anti-disturbance management model of the enterprise supply chain based on the HPSO-R algorithm proposed in the study is better than the comparison model when applied in practice.

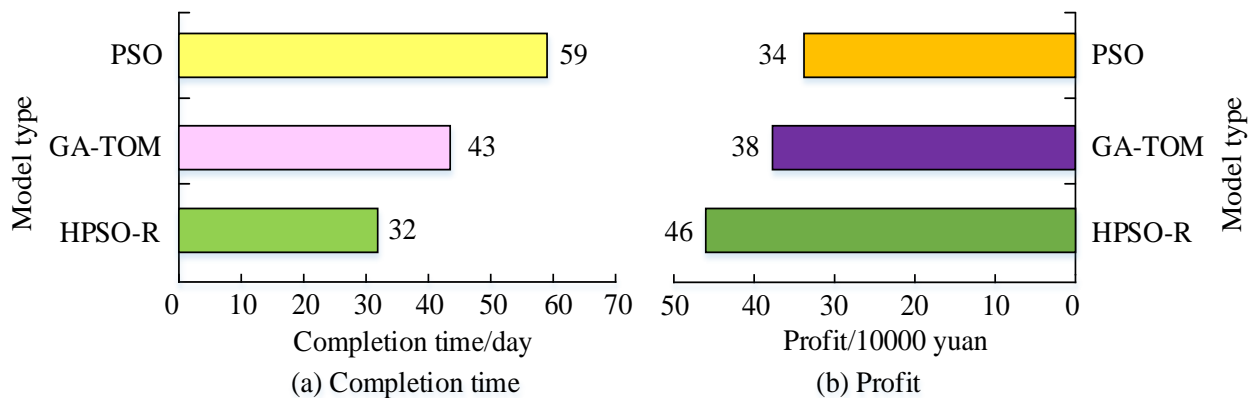


Fig. 10. Time to completion and profit.

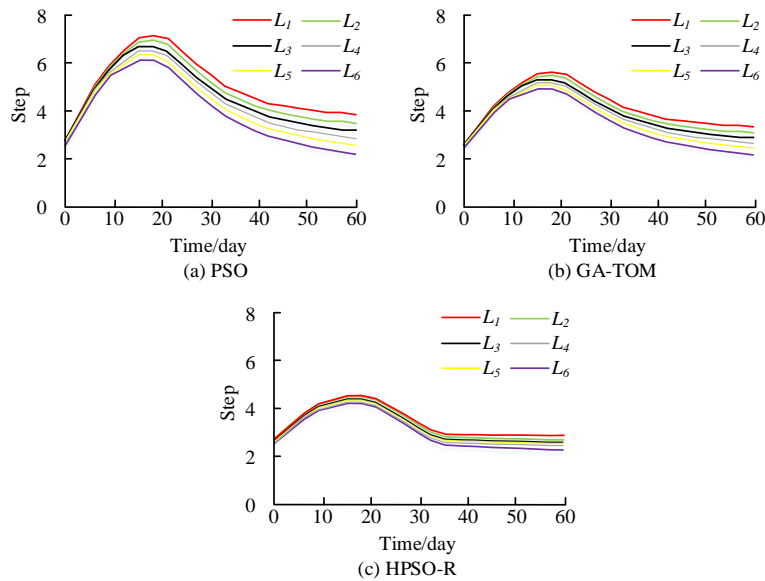


Fig. 11. Comparison of antiproductive behavior.

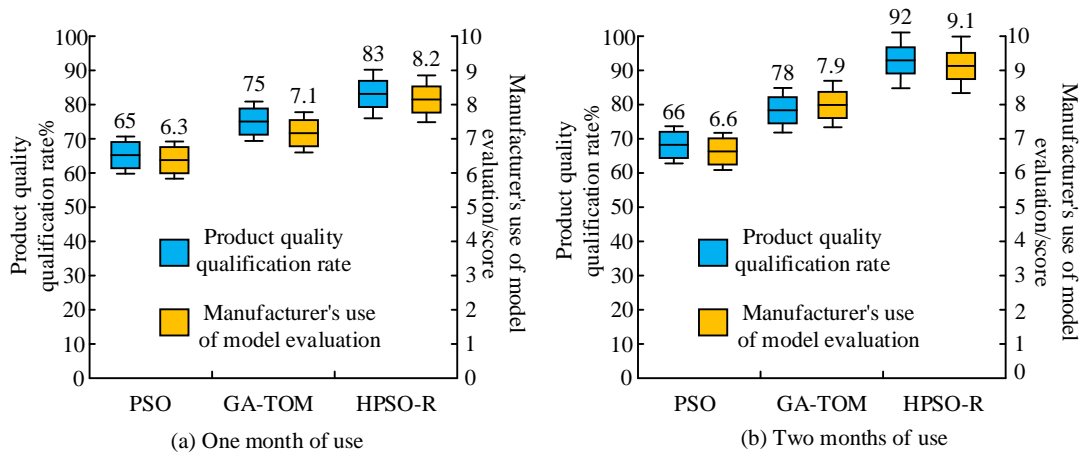


Fig. 12. Pass rate of goods and manufacturer score.

C. Comparative Analysis of Improved Algorithm Performance is discussed

In the comparative test of the performance of the improved algorithm, ONVG, UDNS, NISD, ADF and the particle swarm fitness value of the algorithm are proposed as evaluation indexes. Among them, ONVG, UDNS, NISD and ADF are all evaluation indexes of non-dominated ranking problems. Non-dominant ranking is to divide all non-dominant individuals into the first non-dominant optimal layer, and formulate and assign a shared virtual fitness value. The already stratified individuals are then ignored and the division continues, with a second non-dominant layer appearing, which also develops and assigns a shared virtual fitness value, and so on until all population individuals are stratified. This has the advantage of good individual fitness values and also maintains population diversity. Therefore, it is applied to the performance comparison of multi-objective optimization algorithms. Singh et al. adopted non-dominated ranking index to verify the optimization algorithm of multi-objective problem in the dynamic balance mechanism of cleaning

device of agricultural thresher. The results show that the proposed method is feasible, that is, non-dominated ranking index is universal for the detection of multi-objective optimization algorithm [21]. In the study of the basic method of non-orthogonal multiple access which has been envisaged as the fifth generation cellular network, Kumaresan's team used the fitness value of each iteration of the algorithm to evaluate the particles in the validation of the adaptive user clustering algorithm based on the particle swarm optimization algorithm. The results show that the proposed algorithm has advantages. That is, the particle swarm fitness value algorithm detection of the algorithm is scientific [22]. Therefore, it is feasible to use ONVG, UDNS, NISD, ADF and the particle swarm fitness value of the algorithm as the evaluation index of the algorithm in the experiment. In ten comparison tests, the ONVG values of HPSO-R algorithm are 13, 13, 13, 13, 14, 12, 12, 11, 12 and 10, respectively, which are greater than the other two comparison algorithms, indicating the optimal performance of the algorithm. The minimum UDNS values of HPSO-R algorithm are 1.6 and the maximum UDNS values are 4.2, respectively, which are both lower than the other two

comparison algorithms, indicating that the algorithm has the best performance. The maximum ADF value of HPSO-R algorithm is 5.6, the minimum is 2.4, and the average value is 4.1, all of which are lower than the other two comparison algorithms, indicating that the algorithm has the best performance. The maximum NISD value of HPSO-R algorithm is 0.62, the minimum is 0.11, and the average value is 0.31, all of which are greater than the other two comparison algorithms, indicating that the algorithm has the best performance. The smaller the particle swarm fitness value is, the smaller the gap between the operation result and the optimal solution, that is, the better the algorithm performance. In the iterative test, the particle swarm fitness value of HPSO-R algorithm is finally stable at 0.016, the lowest value, and the fastest convergence speed, which is better than the comparison algorithm. In summary, the HPSO-R algorithm proposed in this study has the advantages of fast running speed and high accuracy, and its performance is better than the comparison algorithm.

V. CONCLUSION

A robust company supply chain is crucial to the successful growth of a business. The production practises of manufacturers have a greater impact on the supply chain than any other of the different links that make it up. The study suggests combining evolutionary algorithms and random neighbourhood structures to improve the particle swarm optimisation method in order to address this issue, and it then completes the building of the enterprise supply chain's anti-disturbance management model using the revised algorithm. When the modified method was tested for comparative performance, it outperformed both the PSO algorithm and the GA-TOM algorithm with maximum ONVG and NISD values of 14 and 0.62, respectively, and minimum UDNS and ADF values of 1.6 and 2.4, respectively. The algorithm outperforms the comparison algorithms in both the quickest convergence rate and the smallest particle swarm fitness value, which is steady at 0.016 after 800 iterations. In a trial of the practical application effects of the model proposed in the study, it was found that the production team using the model had the shortest time and highest profitability in completing the same order volume, 32 days and \$460,000 respectively. Additionally, in the experiment, the members of this group outperformed both the PSO model and the GA-TOM model by having the lowest values for each of the six forms of anti-production behavior—4.5, 4.5, 4.3, 4.3, 4.2, and 4.2, respectively. At one month into the experiment, the group employing this model had a manufacturer score of 8.2 and a merchandise conformance rate of 83%. The model group's pass rate and manufacturer score at the halfway point of the testing were 92% and 9.1, respectively. In conclusion, the study's anti-disturbance management model of the company supply chain may realistically schedule the behaviour of manufacturers with regard to production, hence enhancing the stability of the supply chain.

In the study of the influence of behavioral factors, although the measurement method is obtained, the parameters are selected by direct reference to the values in the existing literature. In the subsequent work, empirical studies can be conducted according to different enterprises and environments,

so as to be closer to the actual situation. In supply chain management, because the member enterprises of the supply chain are independent economic entities, their interests and goals are different, so the coordination and decision-making problem is more complicated and diversified, and it is urgent to study it deeply. At the same time, this model is not strong enough for the connection and coordination between all links in the supply chain. How to do a good job in the connection between all links is also a problem to be further studied in the future.

REFERENCES

- [1] He D, Zhang Z, Han M, Kang Y, Gao P. Multi-dimensional boundary effects and regional economic integration: Evidence from the Yangtze River Economic Belt. *International Regional Science Review*, 2022, 45(4):472-498. DOI:10.1177/01600176211061831.
- [2] Lu Q, Liu B, Song H. How can SMEs acquire supply chain financing: the capabilities and information perspective. *Industrial Management & Data Systems*, 2020, 120(4):784-809. DOI:10.1108/IMDS-02-2019-0072.
- [3] Lau H, Tsang Y, Nakandala D, Lee C. Risk quantification in cold chain management: a federated learning-enabled multi-criteria decision-making methodology. *Industrial Management & Data Systems*, 2021, 121(7):1684-1703. DOI:10.1108/IMDS-04-2020-0199.
- [4] Cavicchi C, Vagnoni E. The role of performance measurement in assessing the contribution of circular economy to the sustainability of a wine value chain. *British food journal*, 2022, 124(5):1551-1568. DOI:10.1108/BFJ-08-2021-0920.
- [5] Liu K, Wang C, Liu L, Xu L. Which group should governmental policies target? Effects of incentive policy for remanufacturing industry. *RAIRO - Operations Research*, 2021, 55(3):1579-1602. DOI:10.1051/ro/2021012.
- [6] Zakian P, Ordoubadi B, Alavi E. Optimal design of steel pipe rack structures using PSO, GWO, and IGWO algorithms. *Advances in Structural Engineering*, 2021, 24(11):2529-2541. DOI:10.1177/13694332211004116.
- [7] Wen S, Hong P, Huang P. Multizone barrel temperature control of the eccentric rotor extrusion process. *Journal of Polymer Engineering*, 2020, 40(3):247-255. DOI:10.1515/polyeng-2019-0315
- [8] Javed S, Ishaque K, Siddique S, Zainal S. A Simple yet Fully Adaptive PSO Algorithm for Global Peak Tracking of Photovoltaic Array Under Partial Shading Conditions. *IEEE Transactions on Industrial Electronics*, 2021, 69(6):5922-5930. DOI:10.1109/TIE.2021.3091921.
- [9] Zouari M, Baklouti N, Sanchez-Medina J, Kammoun H, Ben-Ayed M, Alimi A. PSO-Based Adaptive Hierarchical Interval Type-2 Fuzzy Knowledge Representation System (PSO-AHIT2FKRS) for Travel Route Guidance. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 23(20):804-818. DOI:10.1109/TITS.2020.3016054.
- [10] Liu W, Shi C, Zhu H, Yu H. Wireless Sensor Network Node Localization Algorithm Based on PSO-MA. *Journal of web engineering*, 2021, 20(4):1137-1154. DOI:10.13052/jwe1540-9589.2048.
- [11] Song S, Shi X, Song G, Huq F. Linking digitalization and human capital to shape supply chain integration in omni-channel retailing. *Industrial Management & Data Systems*, 2021, 212(11):2298-2317. DOI:10.1108/IMDS-09-2020-0526.
- [12] Long Q, Tao X, Shi Y, Zhang S. Evolutionary Game Analysis Among Three Green-Sensitive Parties in Green Supply Chains. *IEEE Transactions on Evolutionary Computation*, 2021, 25(3):508-523. DOI:10.1109/TEVC.2021.3052173.
- [13] Ratnakar R, Gupta N, Zhang K, Doorne C, Fesmire J, Dindoruk B, Balakotaiah V. Hydrogen supply chain and challenges in large-scale LH2 storage and transportation. *International journal of hydrogen energy*, 2021, 46(47):24149-24168. DOI:10.1016/j.ijhydene.2021.05.025.
- [14] Lin Q, Zhao Q, Lev B. Influenza vaccine supply chain coordination under uncertain supply and demand. *European Journal of Operational Research*, 2021, 297(3):930-948. DOI:10.1016/j.ijhydene.2021.05.025.

- [15] Mushakhian S, Ouhimmou M, Ronnqvist M. Salvage harvest planning for spruce budworm outbreak using multistage stochastic programming. *Canadian Journal of Forest Research*, 2020, 50(10):953-965. DOI:10.1139/cjfr-2019-0283.
- [16] Li Y, Li X, Zhu D, Yang S. SELF-COMPETITION LEADER-FOLLOWER MULTI-AUV FORMATION CONTROL BASED ON IMPROVED PSO ALGORITHM WITH ENERGY CONSUMPTION ALLOCATION. *International Journal of Robotics & Automation*, 2022, 37(3):288-301. DOI:10.2316/J.2022.206-0563.
- [17] Wang Q, Xi H, Deng F, Cheng M, Buja G. Design and analysis of genetic algorithm and BP neural network based PID control for boost converter applied in renewable power generations. *IET renewable power generation*, 2022, 16(7):1336-1344. DOI:10.1049/rpg2.12320.
- [18] Atanassov K. New topological operator over intuitionistic fuzzy sets. *Journal of Computational and Cognitive Engineering*, 2022, 1(3):94-102. DOI:10.1016/S0165-0114(86)80034-3.
- [19] Tari S, Basseur M, Goffon A. Partial neighborhood local searches. *International Transactions in Operational Research*, 2021, 29(5):2761-2788. DOI:10.1111/itor.12983.
- [20] Lee Y, Su Y, Sun R, Li C. Public responses to employee posts on social media: the effects of message valence, message content, and employer reputation. *Internet Research*, 2020, 31(3):1040-1060. DOI:10.1108/INTR-05-2020-0240.
- [21] Singh P, Chaudhary H. Dynamic balancing of the cleaning unit used in agricultural thresher using a non-dominated sorting Jaya algorithm. *Engineering Computations*, 2020, 37(5):1849-1864. DOI:10.1108/EC-03-2019-0087.
- [22] Kumaresan S, Tan C, Lee C, Ng Y. Low-complexity particle swarm optimisation-based adaptive user clustering for downlink non-orthogonal multiple access deployed for 5G systems. *World Review of Science, Technology and Sustainable Development*, 2022, 18(1):7-19. DOI:10.1504/WRSTSD.2022.119298.

Automated Analysis of Job Market Demands using Large Language Model

Myo Thida* 

Department of Computer Engineering,
Chiang Mai University, Chiang Mai, Thailand

Abstract—This paper presents a comprehensive analysis of labor market demands for Myanmar workers in Japan, and Thailand, focusing on opportunities for individuals without higher education degrees. Leveraging ChatGPT's text classification and summarization capabilities, we extracted vital insights from extensive job advertisements and social media groups. The dataset comprises 152 job advertisements from Thailand and 30 from Japan, collected in 2023. Our research provides a valuable snapshot of skill demands and job opportunities, offering insights for informed decision-making by both job seekers and international non-governmental organizations. The innovative approach of using ChatGPT highlights its efficacy in understanding labor market dynamics. These findings serve as a foundation for tailored interventions to bridge employment challenges faced by marginalized Myanmar youths.

Keywords—ChatGPT; labour market analysis; skills identification; online job adverts; skills demand

I. INTRODUCTION

Understanding the skill demands of industries is crucial for educational providers, graduates, and job seekers. In today's rapidly changing job market, having the right skills is vital for securing desired positions and achieving career success. Educational institutions play a key role in preparing individuals for the workforce. However, without an accurate understanding of industry skill requirements, their programs may not align with market needs, leaving graduates unprepared for employers' expectations.

For educational providers, insights into industry skill demands enable them to design and customize curricula to meet job market needs. By staying updated on evolving skill requirements, institutions can update courses, introduce new programs, and adapt teaching methods to ensure graduates possess sought-after skills. This alignment enhances graduates' employability, improves job prospects, and boosts the institution's reputation.

Likewise, for graduates and job seekers, understanding industry skill demands is equally important. It allows them to make informed decisions about education and training choices. By knowing which skills are in high demand, job seekers can acquire or enhance those skills to improve their competitiveness. Aligning their skill sets with industry requirements increases job offers, provides more opportunities for career advancement, and contributes to long-term success.

The digitization of the job market has created opportunities to better understand job market needs through the accessibility of online job advertisements. These advertisements serve as valuable sources for understanding the most

popular job openings and skill requirements in the market. However, job postings often contain unstructured text and require further processing to identify the required skills. This is where Natural Language Processing (NLP) techniques come into play. NLP has gained significant attention for its ability to identify and extract skills mentioned in job advertisements. Various methodologies and approaches, such as named entity recognition, rule-based systems, machine learning algorithms, and information retrieval models, have been proposed for skill extraction from job advertisements using NLP. Extensive research has been conducted on the extraction of demanded skills from job advertisements, uncovering valuable insights into job market requirements and skill trends.

Recently, large language models have emerged as powerful tools for processing and generating human-like language. Its capabilities stimulates the interests of researchers to leverage this tool in the context of identifying high-demand skills in the dynamic job market. The primary objective of this research is to utilize ChatGPT's capabilities in text classification and summarization to pinpoint the currently sought-after skills. It is crucial to emphasize that this research specifically concentrates on the skills needed by skilled workers without higher formal education, with a particular emphasis on Myanmar youths as a case study. The aim is to assess the employment prospects available in Thailand and Japan for Myanmar youths lacking higher education degrees.

Myanmar youths have been grappling with significant challenges, including limited job opportunities and access to education. The employment rate in Myanmar has experienced a decline of 4.8 percentage points between 2020 and 2022, indicating a decrease in job opportunities [WorldBank, 2023] [1]. Many individuals from rural areas have resorted to seeking work overseas, both legally and illegally, as a means of survival [2]. Unfortunately, their dire circumstances often leave them with limited options and force them to accept any available job to sustain themselves. Consequently, a large portion of these individuals find employment as general workers across various factory sectors, without opportunities to enhance their skills or advance their careers. Prior to the Covid-19 pandemic, it was estimated that around three million Myanmar migrant workers were employed in Thailand. These workers typically found employment in sectors such as fishing, seafood processing, factories, and agriculture. Similarly, the number of Myanmar nationals working in Japan has also been increasing, with more than 33,000 workers reported in 2020. These figures are likely to have grown since then, indicating a growing trend of Myanmar citizens seeking employment opportunities in Japan and Thailand [3].

Traditionally, recruitment for these industries was facilitated through recruitment agencies or community networks, with a focus on individuals who had basic proficiency in the Thai language rather than specific vocational skills. Unfortunately, these workers often face difficulties such as lower pay, workplace abuse, long working hours, and a loss of dignity. In recent years, alongside the continued prevalence of agent-based recruitment, there has been an increasing trend among Myanmar workers to explore alternative channels for finding employment opportunities. This includes searching government and company websites, joining relevant Facebook groups, and utilizing job advertisement platforms. These online avenues provide additional options for job seekers to connect with potential employers and access a broader range of opportunities.

This study aims to provide a snapshot of employment opportunities for Myanmar youths seeking overseas jobs through online channels. It focuses on identifying specific labor market opportunities in Japan and Thailand for Myanmar workers. By gaining insights into job demands and the relevant skills required in these countries, the analysis aims to assist Myanmar job seekers, especially those lacking higher education degrees. To conduct this research, data was collected through job recruitment agencies and by directly collecting data from job advertisements, providing insights into current job opportunities and requirements.

This research makes the following significant contributions:

- This research serves as a proof of concept for the effectiveness of using ChatGPT for skill extraction from job advertisements. By demonstrating the tool's capabilities in extracting relevant skills from unstructured data, it showcases its potential for analyzing job market demands.
- This research uniquely focuses on job prospects for Myanmar youths in Thailand and Japan without higher education. While many studies center on professionals with formal degrees, this research highlights skilled laborers without such qualifications.
- This research highlights challenges faced by Myanmar youths in job and education. It aims to contribute to social development by addressing these issues and promoting inclusive opportunities for marginalized individuals without higher education.
- By understanding the job prospects and skills demanded in Thailand and Japan, individual job seekers can equip themselves with the necessary skills and prepare for better employment opportunities. This information also enable policymakers to design targeted interventions to address employment challenges.

In the next section, a detailed review of the literature in the field of skill identification from online job adverts will be provided. This will be followed by the methodology section, which will explain the data collection methods, data pre-processing, and the results of the exploratory data analysis, presented in Section IV. The key findings and recommendations will be provided in Section V. Finally, the paper will be concluded with a summary and suggestions for future research in the final section.

II. LITERATURE REVIEW

Over the past few years, numerous research papers and surveys have been published, delving into various aspects of job market analysis. One promising research direction focuses on developing skill databases such as ESCO (European Skills-/Competences, qualifications and occupations framework) [4], O*NET (Occupational Information Network) [5] to highlight in-demand skills. ESCO is a project that classifies skills, occupations, and related competencies in various European languages. On the other hand, O*NET, developed and maintained by the US Department of Labor, provides comprehensive information about different occupations, including required skills, knowledge, and work activities. These databases, which are regularly updated to reflect changes in the labor market, serve as valuable resources for understanding industry trends and skill requirements.

The trend of customizing research for specific industries or regional analyses is evident in recent studies. Gröger et al. [6] developed a system that automatically identifies skills in German-language job advertisements, showcasing the effectiveness of this approach. Similarly, Papoutsoglou et al. [7] presented a framework for collecting online job advertisements from StackOverflow and extracting the necessary skills and competencies for specific IT jobs. They employed multivariate statistical data analysis to explore correlations within the dataset. Another relevant study conducted by Kennan et al. [8] analyzed online job advertisements to gain insights into the knowledge, skills, and competencies sought after for early career information systems (IS) graduates in Australia. In addition, Adan et al. [9] introduced C3-IoC, an AI-based solution aimed at assisting students from the UK in exploring IT career paths based on their education level, skills, and prior experience.

A recent publication, [10], presented a systematic review of advancements in skill identification based on job market demands from online job advertisements. The authors thoroughly examined 108 research articles published between 2010 and 2020, providing a comprehensive survey on skill identification. Their study established a framework that addresses three key challenges: skill base generation, skill identification methods, and skill identification granularity. To enhance skill identification and capture the dynamic needs of the job market, the authors recommended leveraging recent deep learning methods.

The field of natural language processing has experienced remarkable progress since the introduction of ChatGPT and its subsequent version, GPT-4 [11]. These advancements have had a significant impact on various conventional tasks, including machine translation, sentiment analysis, text summarization, named entity recognition, and topic segmentation, among others. Researchers have been leveraging the capabilities of ChatGPT to enhance the efficiency and effectiveness of text analysis processes. For example, Hoes et al. [12] investigated the potential of ChatGPT for automated online content moderation. Their study demonstrated that ChatGPT achieved an impressive accuracy of 69% in categorizing statements as true or false.

In a similar vein, [13] proposed AugGPT, a text data augmentation approach that utilizes the capabilities of Chat-

GPT. AugGPT generates multiple conceptually similar yet semantically distinct samples by rephrasing each sentence in the training set. These augmented samples can be effectively utilized for downstream model training. Experimental results on few-shot learning text classification tasks demonstrate the superiority of AugGPT over state-of-the-art text data augmentation methods in terms of testing accuracy and the distribution of augmented samples.

The process of skill identification and normalization from job advertisements encounters numerous challenges, as highlighted in [10]. The language used in job adverts can be diverse and informal, leading to ambiguity and noise during skill extraction. Moreover, the ever-changing nature of job markets and evolving skill requirements pose difficulties in maintaining up-to-date skill taxonomies. Recent research papers [12], [13] showcasing the capabilities of ChatGPT in extracting key information and analyzing text data have motivated us to explore the potential of this large language model in automating skills identification from job advertisements.

III. DATASET

A. Data Collection

The dataset for this research was compiled through manual extraction from two job advertisement platforms, as well as various Facebook groups frequently promoting job listings. The dataset for the Thailand job market originated from job advertisements featured on two prominent Thai job recruitment platforms [14] [15] and multiple Facebook groups [16]–[20] known for posting job vacancies. To narrow the focus on skill requirements for individuals without higher education degrees, job advertisements mandating such qualifications were excluded. The majority of these job listings do not specify gender preferences, implying that the positions are open to applicants of any gender. A total of 152 job advertisements were collected between April 2023 and June 27, 2023, offering a snapshot of the Thailand job market during that specific timeframe.

Regarding the Japanese job market, the job advertisement platforms were exclusively aimed at candidates with higher education degrees. Consequently, the dataset for the Japanese job market was procured from job advertisements published on the official Facebook pages of Myanmar recruitment agencies targeting opportunities in Japan [21]–[23]. Data collection spanned from January 2023 to June 22, 2023, and encompassed a total of 30 job advertisements, advertising a total of 842 job openings.

To ensure alignment with the research focus on comprehending skill demands for individuals without higher education degrees, job advertisements requiring higher education degrees or above were excluded during the data collection process. In sum, a total of 152 job advertisements were amassed from April 2023 to June 27, 2023, providing a snapshot of the job market within that specific time frame.

The job advertisements collected were initially in an unstructured format, as shown in Fig. 1. To transform the unstructured content into a structured dataset, the data collectors manually extracted the ‘Job Titles’, and ‘Skills and Responsibilities’. This involved organizing the content from the advertisements into a structured table format, exemplified

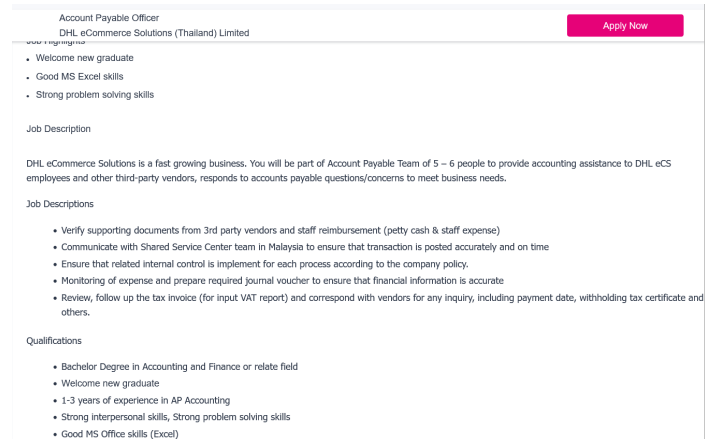


Fig. 1. Example job adverts from JobDB [14].

in Table I. Notably, job titles and skills are often not explicitly mentioned in the job adverts. Consequently, during the manual extraction process, the data collectors copied and pasted the content of the job descriptions or responsibilities into the skill and responsibility column, without explicitly assigning them to pre-defined skill databases like O*NET.

It is important to note that web scraping techniques were not employed in the data collection process due to strict prohibitions by the job advertisement platforms. To ensure compliance with legal regulations and respect the privacy policies of the companies, the researchers collected and categorized the data manually.

The dataset consists of two attributes: ‘Job Titles’, and ‘Skills and Responsibilities’. Table I presents a snapshot of the dataset, showcasing specific details and examples of the collected data.

B. Data Pre-processing

The initial step in data pre-processing for this project involves cleaning the unstructured text data, as presented in Table I. The data contains unwanted symbols, text, and duplicated records, requiring necessary cleansing. Consequently, the data cleaning process encompasses three essential actions: removing white spaces, eliminating duplicates, and converting the text to lowercase. These steps ensure text standardization and facilitate subsequent analysis.

However, it is important to note that the job titles and skills are preserved in their raw form without lemmatization. This deliberate decision was made to evaluate the performance of ChatGPT in grouping similar words with the same meaning. By retaining the original form of job titles and descriptions, we can assess the model’s ability to recognize and interpret variations in language while comprehending contextual information.

IV. PROPOSED METHOD

Our research introduces a novel method that leverages ChatGPT, a large language model, to enhance the efficiency of automated job market analysis and gain insights into job

TABLE I. SNAP-SHOT OF THE COLLECTED DATASET

Job title	Skills and Responsibilities
Service Technician (Air Conditioning)	Respond to customer requests for repair of air conditioning systems in homes companies and factories. Perform routine maintenance tasks to ensure air conditioning systems are operating efficiently. Manage a team of technicians providing guidance and support as needed.
store staff	Responsible to Store Performance and staff development to meet business objective. Maintain and develop retail Store daily Operation to meet business efficiency in the best possibility way. Manage and retain people to run daily operation to achieve customer target and customer satisfaction. Develop SOP and Daily Operation Routine (WI)
Graphic Designer	”Create and conceptualize where needed marketing communications in different formats including advertising social media content website banners video editing (and occasional filming) event collateral and more. Implement visual marketing collateral according to BayWa brand guidelines. Coordinate with the regional marketing and sales teams to create communications that meet their needs. Manage and oversee third party suppliers
Accounting officer	Must be proficient in the use of the English language (both verbal and written) Must be proficient in Thai (both verbal and written) knowledge of other languages is an advantage Excellent organizational and time-management skills Must be very detail-orientated Positive attitude and sincere desire to learn on the job

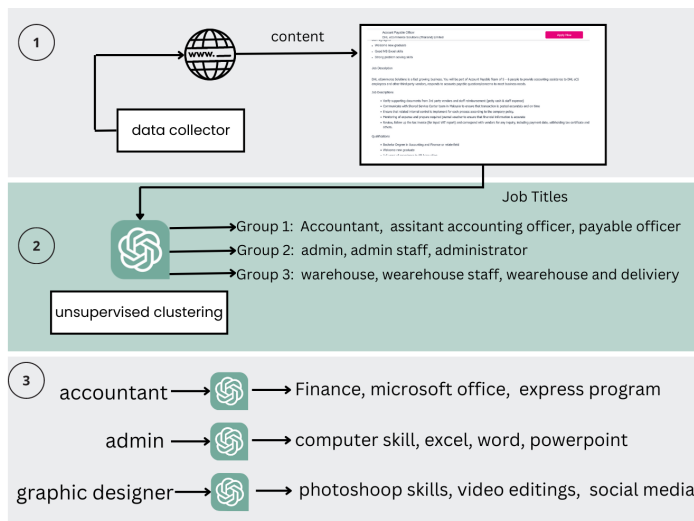


Fig. 2. Architecture of our proposed method.

market demands. The architecture of our proposed method is illustrated in Fig. 2.

The proposed method consists of two key steps aimed at analyzing job titles and skill demand using ChatGPT’s zero-shot prompt capability.

In the initial step, we leverage ChatGPT’s zero-shot prompt capability to perform job title clustering in a given dataset, obviating the necessity for predefined labels or manual annotation. Through the formulation of specific prompts, we guide the model to acquire knowledge from the provided information and infer appropriate clusters based on the semantic relationships among job titles. By applying this clustering process and subsequently analyzing the size of each resultant cluster, we can discern the most sought-after job titles within the market.

The maximum number of clusters can be predetermined within the prompt by specifying either the desired number of clusters or the minimum size of each cluster.

Building upon the outcomes obtained from the Job Title Clustering step, we progress to the subsequent step, wherein we undertake an analysis of skill demand within each cluster. To accomplish this, we rely upon ChatGPT’s zero-shot prompt functionality. For each job title cluster derived from Step 1, we construct a prompt that facilitates the extraction of skills from job descriptions specifically associated with that particular cluster. By employing this approach, we gain valuable insights into the skill demand prevalent within each distinct job title group.

A. Step 1: Job Title Clustering

In the initial step, we leverage the zero-shot prompt capability of ChatGPT to cluster the job titles in the given dataset. This prompt serves as a guide, enabling ChatGPT to generate meaningful clusters based on the similarities; it recognizes among the job titles. The resulting clusters offer valuable insights into different job categories within the job market, facilitating a comprehensive analysis of job market demands. Some job roles appeared multiple times in the job adverts database, indicating a higher demand for those specific roles during the analyzed time frame. Conversely, certain roles occurred rarely, suggesting a relatively lower demand for those positions. Analyzing the sizes of these clusters, in terms of the number of job advertisements assigned to each role, provides valuable insights into the demand for specific job roles during the specified time frame. Larger clusters indicate highly sought-after job roles, implying a higher demand in the job market for roles with the associated skills and qualifications.

Using the following prompt, we guide ChatGPT to recognize underlying patterns and similarities among the job titles. In this method, the number of clusters is not limited, but

we predefined the minimum cluster size as 3, considering the content of the dataset.

```
def job_title_clustering(Job_title):
    prompt = f''''''
    Your task is to group Job titles from \
    a long job title corpus where each job \
    title is separated by comma.

    From the given job title corpus, delimited \
    by triple quotes \
    group the job titles that are similar.

    The minimum size of the group should be 3.

    Use the following format:
    Group 1: [<list of job titles that are \
    similar>, quote each job title with ' ' \
    ]
    Group 2: [<list of job titles that are \
    similar>, quote each job title with ' ' \
    ]

    ```{Job_titles}```
    ```
    response = get_completion(prompt)
    return response
```

Listing 1: Prompt for Job Title Clustering

Table II presented as a snapshot of the clustering results, showcases the various job roles obtained from the clustering process in Step 1.

TABLE II. SNAP-SHOT OF THE JOB ROLES FROM THE CLUSTERING PROCESS

Group 1	Group 2	Group 3	Group 4
account payable officer	customer relations officer	administrator	warehouse
accountant	customer service	admin staff	warehouse and delivery staff
accountant	customer service	admin	warehouse associate
accounting and finance officer	customer service assistant (english speaking)	admin account	warehouse operations
accounting officer (ap)	customer service officer	admin executive	warehouse staff
assistant accounting manager	customer service quality assurance	admin officer	

B. Step 2: Skills Extraction

In the second step, we utilize ChatGPT’s zero-shot prompt capability to extract skills from each cluster. To accomplish this, we gather the job descriptions from all the job advertisements within the corresponding cluster and combine them

into a comprehensive string. This concatenated job description corpus is then inputted into the model, prompting it to extract skills from the descriptions.

Below is a code snippet illustrating the prompt employed to extract skills from a job description:

```
def extract_skills(grp_skill):
    prompt = f''''''
    Your task is to extract skills from \
    a corpus of job descriptions.

    From the given corpus of job descriptions, \
    delimited by triple quotes \
    extract the commonly found skills.

    Format your response as a list of skills \
    separated by comma.

    ```{grp_skill}```
    ```
    response = get_completion(prompt)
    return response
```

Listing 2: Prompt for Skills Extraction

This prompt aids in guiding the model to identify and extract relevant skills from the job descriptions, contributing to a more comprehensive understanding of the skill requirements within each cluster. The *extract_skills* function takes a parameter called *grp_skill*, which represents the concatenated job descriptions from each cluster. By utilizing this prompt, it assists in guiding the model to identify and extract pertinent skills from the job descriptions. This process significantly contributes to enhancing the overall understanding of the skill requirements associated with each job role within the clusters. Through this step, we can effectively identify the key skills needed for each job role and gain valuable insights into the skill demand across different clusters.

V. RESULTS

A. Top Demanded Job Titles

The task of clustering job advertisements into common job roles, such as ‘accountant’ or ‘admin’, poses a significant challenge. The absence of standardized definitions for these roles and the existence of multiple names for the same role make the task non-trivial. Furthermore, not all job advertisements can be easily categorized into predefined roles. In order to address these difficulties, we leveraged ChatGPT’s zero-shot prompt to cluster job titles and identify common job roles. We extracted the ten job roles with the largest number of job advertisements.

1) *Thailand*: Fig. 3 presents the top ten job positions that were highly sought after by low-level skilled workers in the Thailand job market between April and June 2023. Sales-related roles accounted for approximately 15% of the job advertisements during this period, including job titles such as *Sales*, *Sales Admin/Support Executive*, and *Sales Coordinator*. The second most in-demand category was customer service-related positions, representing around 13% of the job advertisements. This category encompassed roles like *Customer Relations Officer*, *Customer Support Specialist*, and *Customer*

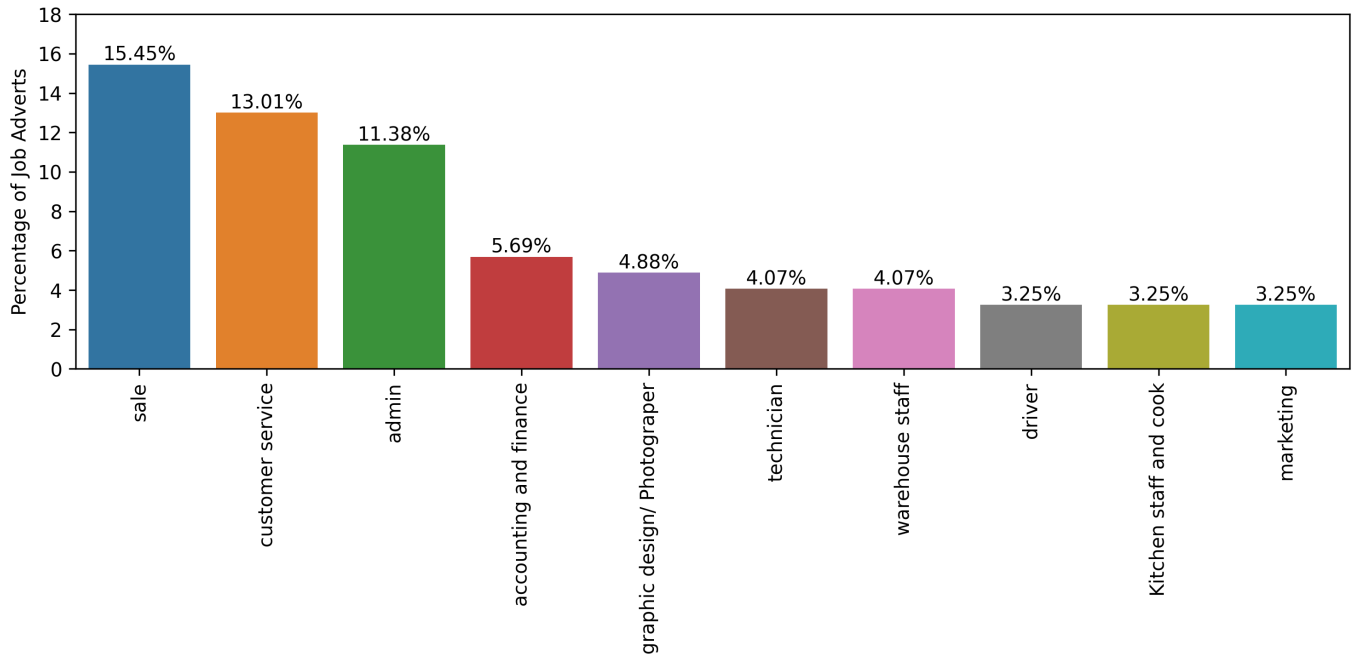


Fig. 3. Top ten most in-demand job positions in Thailand job market.

Service (Call Center). Admin-related positions ranked third in terms of demand and included roles such as *Admin Executive*, *Admin Officer*, *Admin Support Staff*, and *Administrative Officer*. It is observed that all the job postings are open to both male and female candidates.

TABLE III. COMPARISON OF JOB ROLES FROM HUMAN ANNOTATION AND CHATGPT

ChatGPT	%(GPT)	Manual	%(Manual)
sale	15.45	sales	12.67
customer service	13.01	customer service	12.67
admin	11.38	admin	12.0
accounting and finance	5.69	marketing	7.33
graphic design/ Photographer	4.88	staff	5.33
technician	4.07	accountant	5.33
warehouse staff	4.07	warehouse	4.67
driver	3.25	receptionist	4.67
Kitchen staff and cook	3.25	officer	4.67
marketing	3.25	technician	4.0

Comparison against human annotations: To assess the clustering results, we compared the top ten job titles obtained through our proposed method with the annotations provided by job-seeking individuals in the Thailand market. The comparison results are shown in Table III, which illustrates the overlapping job roles between the individuals' specifications and the results generated by ChatGPT. The frequency column

represents the proportion of job advertisements related to specific job roles within the analyzed time-frame. Our observations revealed that seven out of the top 10 job roles, excluding graphic designer, driver, and kitchen staff, overlapped between the human annotations and the model's results. The top three job roles remained consistent, with only minor variations in the frequency of job advertisements.

2) *Japan:* Fig. 4 illustrates the top ten job positions that were in high demand among low-level skilled workers from Myanmar in the Japanese job market between January and June 2023. The analysis considered 30 job advertisements, totaling 842 job openings. Please note that for the Japanese Job market, worker recruitment in Myanmar is commonly facilitated through agencies, and in many cases, mass recruitment is conducted where more than 10 workers are hired simultaneously.

The top ten industries with high demand encompassed agriculture, food service, interior building cleaning, and nursing care. Among these industries, agriculture recorded the highest level of recruitment, employing over 250 workers (30% of total job openings). The food service industry closely followed, hiring approximately 120 workers. The cleaning industry had the highest recruitment of female workers, with 100 positions filled during the study period. The nursing care sector and fishery/aquaculture industries each recruited around 65 workers. Additionally, the manufacturing of food and beverages industry hired approximately 50 workers.

Out of all the job advertisements, 14% specifically targeted female workers, including the cleaning industry, manufacturing of food and beverages, and airport ground staff. Conversely, only 10% of the job adverts were specifically aimed at male workers. Notably, the construction industry exclusively recruited male workers, with a total of 30 workers being hired.

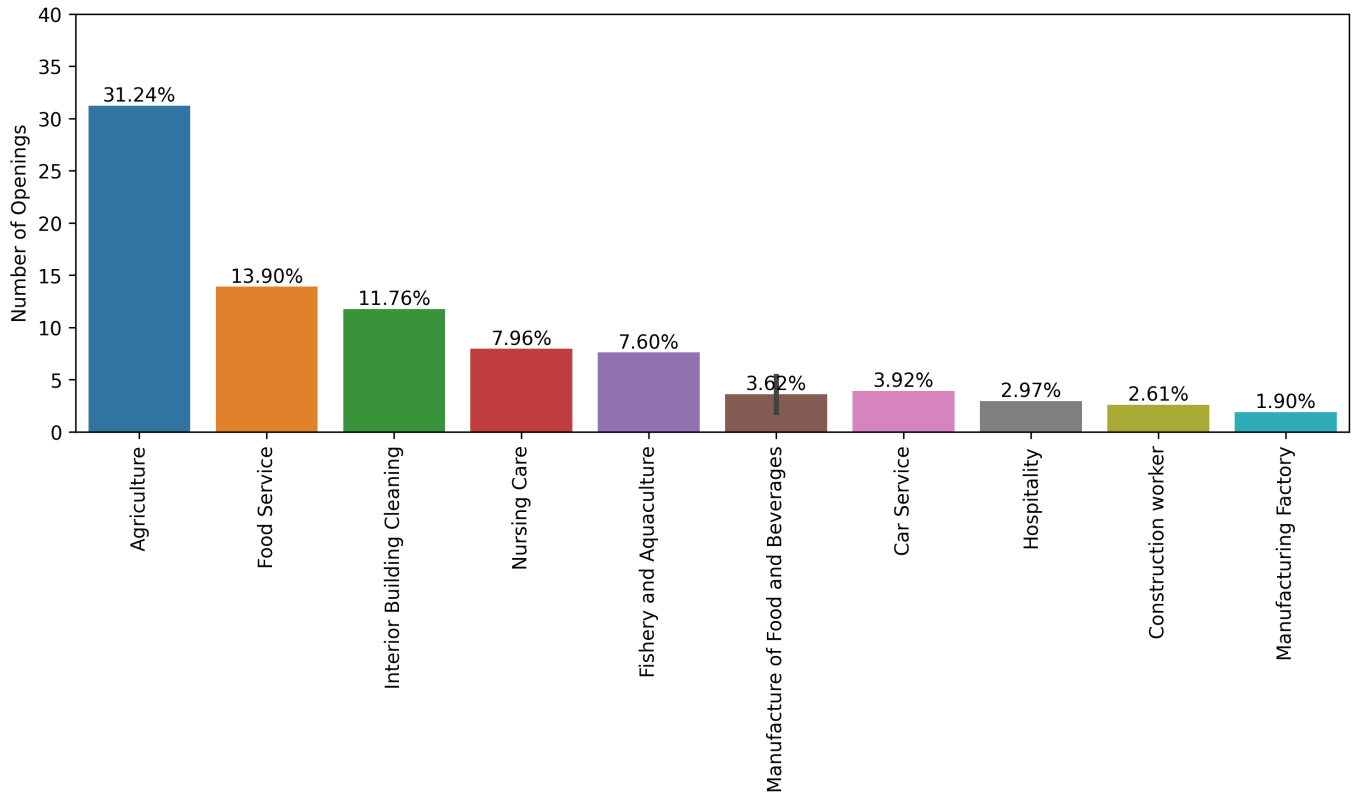


Fig. 4. Most in-demand job sectors in Japanese job market targeting to Myanmar skilled workers.

The carpentry and interior home decoration industries followed as the second and third most recruited industries for male workers. The majority of job recruitment, accounting for 76%, was open to both male and female workers. Additionally, it is worth mentioning that while some openings in the manufacturing of food and beverages industry were exclusively for female workers, others were open to both male and female applicants.

Comparison against Human annotations: Table IV presents a comparison of the results obtained from Human annotations and those generated by ChatGPT. The table reveals a significant level of overlap in clustering between the human annotations and the model's output, with only minor variations in the frequency of job advertisements.

3) *Comparison between Thailand and Japan:* Please note that the above data represents two different dimensions, providing insights into the job markets in Thailand and Japan for Myanmar workers. The data from Thailand encompasses job advertisements that are open to not only Myanmar nationals but also individuals of other nationalities, including Myanmar workers. On the other hand, the Japanese data specifically focuses on opportunities for Myanmar workers in Japan.

Both Japan and Thailand utilize their national languages as the official working languages. However, in the Thai job market, there is a noticeable trend of posting job advertisements in English. This practice aims to attract individuals of diverse nationalities, including Myanmar workers who possess English language skills. On the other hand, the Japanese job market follows a different approach. Job advertisements in English primarily target professionals with higher degrees or

TABLE IV. COMPARISON OF JOB ROLES FROM HUMAN ANNOTATION AND CHATGPT

ChatGPT	%(GPT)	Manual	% (Manual)
Agriculture	31.24	Agriculture workers	31.97
Food Service	13.9	Food service	14.63
Interior Building Cleaning	11.76	Cleaner	12.49
Nursing Care	7.96	Nurse and care giver	8.69
Fishery and Aquaculture	7.6	Fishery Firm workers	8.34
Manufacture of Food and Beverages	5.34	Construction worker	6.91
Car Service	3.92	Manufacture of Food and Beverages	6.08
Hospitality	2.97	Car Service	4.66
Construction worker	2.61	Manufacturing Factory worker	4.06
Manufacturing Factory	1.9	Front Desk	3.71

specialized skills. These advertisements cater to individuals who possess a certain level of proficiency in English, reflecting the demand for language fluency in certain industries or job

roles. However, lower-skilled workers from Myanmar seeking employment in Japan often heavily rely on recruitment agencies to secure job opportunities. These agencies act as intermediaries, connecting job seekers with employers who specifically seek foreign workers. In Myanmar, these agencies frequently use social media platforms like Facebook to announce job advertisements, enabling job seekers in Myanmar to search for opportunities through these platforms.

It is important to note that both Thailand and Japan recognize the importance of language proficiency, and fluency in the respective national languages is highly valued. While the Thai job market embraces the use of English in job advertisements to attract a diverse talent pool, the Japanese job market focuses more on higher-skilled positions and relies on recruitment agencies for lower-skilled job placements. Understanding these dynamics is crucial for Myanmar workers seeking employment opportunities in these two countries.

By comparing the study between the Japanese and Thai job markets, it can be observed that the top five industries with high demand for both genders in the Japanese job market were agriculture (31%), food service (14%), building cleaning (12%), nursing care (8%), and fishery/aquaculture. In the Japanese job market, a significant majority of job recruitment (76%) was open to both male and female workers. The construction industry exclusively recruited male workers, while the cleaning industry had the highest recruitment of female workers. In contrast, data from official job agencies indicates that the Thai job market had a high demand for sales-related roles (15%), customer service-related positions (13%), and administrative-related positions (11%) between April and June 2023. This suggests that the service sector in Thailand is becoming more open to Myanmar nationals who can speak both English and Thai.

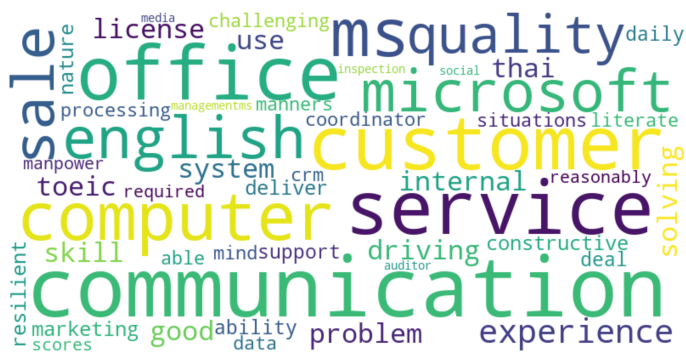


Fig. 5. Snap-shot of the top demanded skills for top three sought-after job roles in Thailand.

B. Top Demanded Skills

Proficiency in Japanese, and Thai is essential for workers seeking promising job opportunities in various sectors.

1) *Thailand:* Fig. 5 displays the required skills for the top three sought-after job roles: sales, customer service, and administration. Among the technical skills, employers highly prioritize computer literacy proficiency and competence in Microsoft Office tools. English language proficiency is also frequently requested. It is worth noting that the Thai job market

places great importance on soft skills as well. Employers in Thailand consistently seek abilities such as customer service, communication, and problem-solving.

2) *Japan:* Fig. 6 provides valuable insights into the required skills for the agriculture and fishery/aquaculture sectors in Japan. In addition to the essential manual labor skills, such as breeding, collecting, sorting of animals, farming, and planting, there is also a significant demand for knowledge in management and health and safety practices. It is worth noting that the agriculture and fishery/aquaculture sectors in Japan require workers with a diverse skill set. While manual labor skills are foundational, the demand for additional knowledge in management and health and safety reflects the need for workers who can contribute to the overall success and sustainability of these industries.



Fig. 6. Snap-shot of the top demanded skills for agriculture and fishery / aquaculture sectors in Japan.



Fig. 7. Snap-shot of the top demanded skills for construction sector in Japan.

Fig. 7 highlights the highly demanded skills for male workers in the Japanese construction sector. It can be seen that the Japanese construction sector demands a range of skills related to various construction activities such as plumbing, pumping, formwork, finishing, scaffolding, plastering, roofing and carpentry. In addition, it also requires skilled workers who can operate and maintain a wide range of machinery and equipment. By focusing on acquiring these skills, Myanmar workers can enhance their chances of finding employment opportunities in the Japanese construction sector.

C. Evaluation of Unsupervised Clustering

To assess the clustering results between the proposed method and manual grouping, we employ precision and recall as evaluation metrics. Precision quantifies the proportion of job roles accurately classified by the model, while recall measures the model’s capability to correctly capture all job titles.

1) *Evaluation metrics:* Precision, as a metric, gauges the correctness of the clustering results by calculating the percentage of job roles that were correctly assigned to their respective clusters. A higher precision value indicates a higher accuracy in classifying job roles.

$$\text{precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}} \quad (1)$$

On the other hand, recall assesses the model’s completeness in capturing all relevant job titles. It measures the proportion of actual job titles that were correctly identified and included in the dataset. A higher recall value implies that the model successfully captures a larger portion of the job titles.

$$\text{recall} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (2)$$

In Eq. (1) and (2), True Positives indicate the number of job titles that are accurately classified within their respective clusters. This means that a job title is correctly assigned to the appropriate cluster, such as “admin officer” being assigned to the admin cluster. False Positives, on the other hand, refer to job titles that are incorrectly assigned to a cluster where they don’t belong. For instance, if a job title like “sale” is incorrectly assigned to the admin cluster. False Negatives represent the job titles that should be belong to the given cluster but are mistakenly assigned to the wrong cluster. For example, the title “admin officer” is erroneously assigned to the sale cluster instead of the admin cluster.

By considering both precision and recall, we can gain a comprehensive understanding of the performance of the proposed clustering method compared to the manual grouping. These evaluation metrics provide insights into the model’s accuracy, correctness, and ability to capture a broad range of job titles within the clustering process.

2) *Accuracy:* In this section, we present the performance of unsupervised clustering for the top ten job roles in Thailand Job Market. The resulting confusion matrix (Fig. 8) provides insights into the clustering accuracy. The confusion matrix visually represents the performance of the clustering model, showcasing its ability to assign job adverts to the corresponding job roles. The high accuracy scores indicate the model’s proficiency in recognizing and grouping similar job descriptions together.

For the Thailand Job market, the proposed model demonstrated impressive performance by accurately capturing all the job adverts annotated by humans in the driver, graphic designer, and kitchen helper roles. Furthermore, it achieved a high accuracy rate of 95% for sales roles and 88% for accountant roles. For sales, administration, and customer service roles, ChatGPT achieved approximately 70% accuracy.

However, ChatGPT did not perform well for the marketing and technician roles due to the wide and diverse range of words used to describe these roles. The inherent variability and ambiguity in job descriptions related to marketing and technician positions pose challenges for accurate clustering.

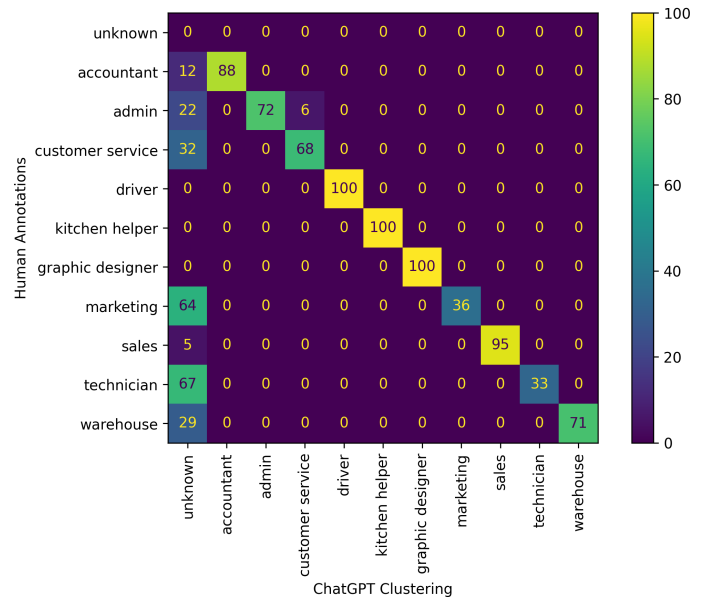


Fig. 8. Confusion Matrix for the top-ten job roles in Thailand.

This evaluation demonstrates the effectiveness of ChatGPT in automatically clustering job adverts into relevant job roles. The results highlight its potential for assisting in job matching and recruitment processes, aiding in the efficient categorization of job postings based on their role requirements.

3) *Precision and recall:* Fig. 9 depicts the precision and recall scores for the top ten demanded job roles in Thailand. The precision scores reflect a high success rate in accurately identifying the job roles, with most of them achieving a score of 100%. This indicates that the model performs well in correctly assigning job postings to their respective roles.

However, the variation in the recall scores indicates the model’s ability to capture all the job roles within the relevant clusters. While precision measures the accuracy of the identified job roles, recall measures the model’s capability to capture all the job roles present in the data. The variation in recall scores suggests that the model may have challenges in fully capturing all job roles, potentially missing some relevant postings.

VI. LIMITATION AND FUTURE DIRECTION

It is important to acknowledge that this research offers a snapshot of the skills in demand by employers at a specific moment, and it may not encompass the complete range of skills required for a given occupation or industry due to the limited dataset size. However, the demonstrated success of the proposed method serves as a promising direction for exploring and harnessing the capabilities of large language models in understanding the industrial skill demands. It provides an avenue to develop a comprehensive understanding of skill

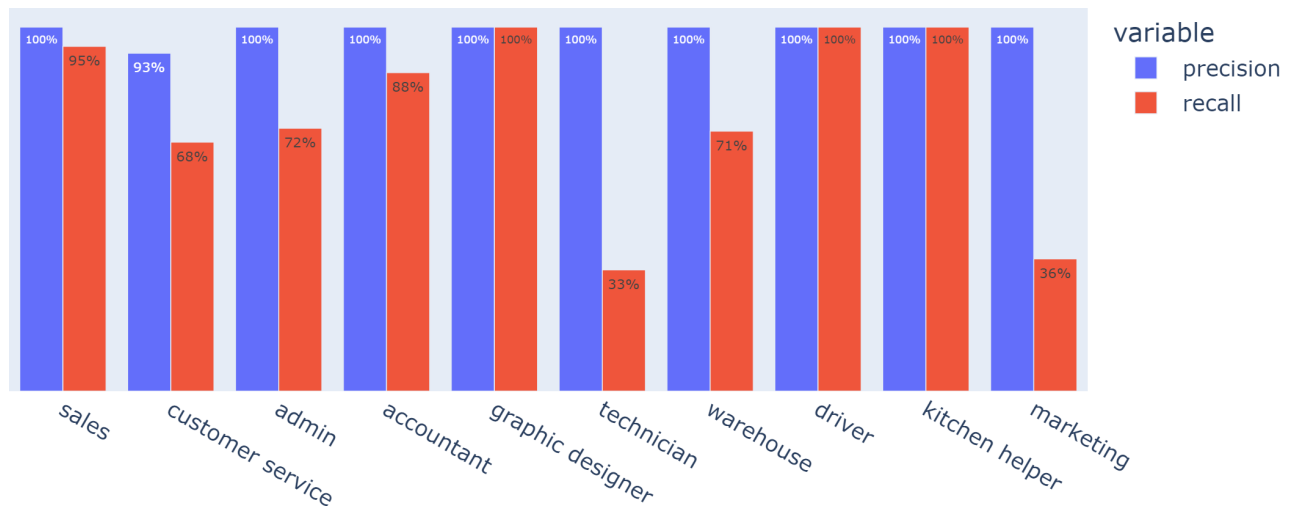


Fig. 9. Precision and Recall for the top-ten job roles in Thailand.

demand and contributes to the advancement of reducing the skill gap between the industries and training providers.

Another limitation is the validation of our results. We engaged youths seeking employment in Thailand and Japan, who lack expertise in labor market studies. This could introduce biased judgments when identifying job titles and clustering data. To improve accuracy, future research could involve experts or existing skill databases, refining job title identification and minimizing biases. This enhances result reliability.

VII. CONCLUSION

This study leverages the capabilities of ChatGPT, a powerful instrument for text classification and summarization, to identify in-demand skills among Myanmar workers. The main objective is to assess employment opportunities in Thailand and Japan for Myanmar youths lacking higher education degrees. Key highlights of this research include showcasing ChatGPT's proficiency in extracting skills from unstructured job ads, offering a rapid and thorough perspective on labor market demand, with a specific emphasis on highlighting opportunities for low-level skilled workers. Moreover, this study empowers international non-governmental organizations to make well-informed decisions while crafting targeted interventions to address the employment challenges confronted by marginalized Myanmar youths.

ACKNOWLEDGMENT

The authors would like to thank the researchers and data analysts from Women in Tech (Myanmar) for collecting the job advertisements manually.

REFERENCES

- [1] S. R. Sutirtha, D. Giorgia, and R. Elizabeth, "A growing crisis: Work, workers and wellbeing in myanmar?" World Bank, Tech. Rep., 2023.
- [2] Z. Nikoloski and R. Smith-Govoni, "Covid-19, coup d'etat and poverty," United Nations(UN), Tech. Rep., 2023.
- [3] G. N. L. of Myanmar, "Youths keen on going overseas for work," May 2022.
- [4] J. Smedt, M. Vrang, and A. Papantoniou, "Esco: Towards a semantic web for the european labor market," vol. 1409, 01 2015.
- [5] "O*net 27.3 database," 2023, <https://www.onetcenter.org/database.html>.
- [6] J. Gröger and G. J. Schneider, "Automated analysis of job requirements for computer scientists in online job advertisements," 2019, p. 226 – 233.
- [7] P. Maria, M. Nikolaos, and A. Lefteris, "Mining people analytics from stackoverflow job advertisements," in *2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, 2017, pp. 108–115.
- [8] M. A. Kennan, D. Cecez-Kecmanovic, P. Willard, and C. S. Wilson, "Is knowledge and skills sought by employers: A content analysis of australian is early career online job advertisements," vol. 15, no. 2, p. 169 – 190, 2009.
- [9] A. José-García, A. Sneyd, and et al., "C3-ioc: A career guidance system for assessing student skills using machine learning and network visualisation," *International Journal of Artificial Intelligence in Education*, pp. 1–28, 2022.
- [10] I. Khaouja, I. Kassou, and M. Ghogho, "A survey on skill identification from online job ads," *IEEE Access*, vol. PP, pp. 1–1, 08 2021.
- [11] "Open ai: Gpt-4," 2023, <https://openai.com/product/gpt-4>.
- [12] E. Hoes, S. Altay, and J. Bermeo, "Leveraging chatgpt for efficient fact-checking," 2023.
- [13] H. Dai, Z. Liu, W. Liao, and et al., "Auggpt: Leveraging chatgpt for text data augmentation," 2023.
- [14] "Jobsdb," 2023, <https://th.jobsdb.com/th>.
- [15] "Jobbkk," 2023, <https://www.jobbkk.com/>.
- [16] "Job postings for thailand industries, facebook group 1," 2023, <https://www.facebook.com/groups/3517244485171573>.
- [17] "Job postings for thailand industries, facebook group 2," 2022, <https://www.facebook.com/groups/565407345189563>.
- [18] "Job postings for thailand industries, facebook group 3," 2023, <https://www.facebook.com/groups/417434729384306/>.
- [19] "Job postings for migrant workers in thailand, facebook group 4," 2020, <https://www.facebook.com/groups/1764364050271625/>.
- [20] "Job postings for migrant workers in thailand, facebook group 5," 2020, <https://www.facebook.com/groups/601115987172804/>.
- [21] "Job recruitment agency - 1, 2023," 2023, [urlhttps://www.facebook.com/Polestar.Employment.Agency/](https://www.facebook.com/Polestar.Employment.Agency/).
- [22] "Job recruitment agency - 2, 2023," 2022, <https://www.facebook.com/profile.php?id=100084530810886>.
- [23] "Job recruitment agency - 3, 2023," 2023, <https://www.facebook.com/popoang.company.limited>.

Decentralized Management of Medical Test Results Utilizing Blockchain, Smart Contracts, and NFTs

Quy T. L.¹, Khanh H. V.*¹, Huong H. L.¹, Khiem H. G.¹, Phuc T. N.¹, Ngan N. T. K.²,
Triet M. N.¹, Bang L. K.¹, Trong D. P. N.¹, Hieu M. D.¹, Bao Q. T.¹, and Khoa D. T.¹

¹FPT University, Can Tho City, Vietnam

²FPT Polytechnic, Can Tho city, Vietnam

Abstract— In today’s medical landscape, the effective management and availability of diagnostic data, including current and historical medical tests, play a critical role in informing physicians’ therapeutic decisions. However, the conventional centralized storage system presents a significant impediment, particularly when patients switch healthcare providers. Given the sensitive nature of medical data, retrieving this information from a different healthcare facility can be fraught with challenges. While decentralized storage models using blockchain and smart contracts have been suggested as potential solutions, these methodologies often expose sensitive personal information due to the inherently open nature of data on the blockchain. Addressing these challenges, we present an innovative approach integrating Non-Fungible Tokens (NFTs) to facilitate the creation and sharing of medical document sets based on test results within a medical environment. This novel approach effectively balances data accessibility and security, introducing four key contributions: (a) We introduce a mechanism for sharing medical test results while preserving data privacy. (b) We offer a model for generating certified, NFT-based document sets that encapsulate these results. (c) We provide a proof-of-concept reflecting the proposed model’s functionality and (d) We deploy this proof-of-concept across four EVM-supported platforms—BNB Smart Chain, Fantom, Polygon, and Celo—to identify the most compatible platform for our proposed model. Our work underscores the potential of blockchain, smart contracts, and NFTs to revolutionize medical data management, demonstrating a practical solution to the challenges posed by centralized storage systems.

Keywords—Medical test result; blockchain; smart contract; NFT; Ethereum; Fantom; Polygon; Binance Smart Chain

I. INTRODUCTION

Advancements in technology are significantly transforming the landscape of disease diagnosis and treatment, alleviating the need for patients to physically visit healthcare facilities. Innovative applications installed on smartphones now enable remote health monitoring, overseen by either human doctors or AI platforms [1]. However, to supplant the entire traditional healthcare system, certain critical steps outlined in numerous research directions must be taken [2], [3], [4]. One significant challenge is the management of individual medical data, including treatment records and medical history. Accurate recording of medical history is vital for effective disease diagnosis and treatment [5].

Many research studies leverage modern technologies to reform healthcare systems, replacing traditional supply chain processes [6], [7], and the way diseases are diagnosed and

treated. These proposed solutions primarily focus on decentralized (or distributed) storage, ensuring efficient data handling and access [8]. Blockchain technology and smart contracts contribute to transparency in information storage [9]. After authentication, all data is stored on-chain and becomes immutable.

Smart contract technology, introduced first by the Ethereum platform¹, automates all system operations. After calculations and updates, information and data are stored on a distributed ledger, accessible for stakeholders to check activities.

Blockchain-based solutions have been proposed to ensure data authentication transparency in the medical environment, addressing shipping [10], [7], disease treatment [11], [12], medical waste management [13], emergency patient information retrieval [14], [15], medical product supply chain [16], and blood donation processes and their supply chain management [17], [18]. Other non-medical solutions based on community sourcing include Cash on Delivery [19], [20], supply chain [21], [22], among others [6], [23].

Several models for managing patients’ medical examination and treatment information based on Blockchain technology have been proposed. For instance, HealthBank² introduced a patient information management model based on Blockchain technology, where users can store all information reliably on the blockchain. Similarly, HealthNautica and Factom Announce Partnership³ utilized the transparency of on-chain storage to build a system protecting medical data integrity.

However, on-chain storage of all patient information encounters two significant issues: i) a decrease in system performance and increased transaction fees per access; ii) lack of patient privacy due to public access to all information. The first problem arises from redundant and unnecessary data storage [24]. Thanh et al. [20] posited that not all collected data needs to be stored and processed on-chain as most are redundant. Trieu et al. [14] shared a similar sentiment, suggesting off-chain storage for personal data unrelated to treatment or diagnosis. This decrease in on-chain data consequently reduces transaction costs [25].

Privacy risk is another issue; unencrypted stored data can be exploited and manipulated by other system users, severely

¹<https://ethereum.org/en/whitepaper/>

²<https://www.healthbank.coop>

³<https://www.factom.com/company-updates/healthnautica-factom-announce-partnership/>

impacting patient privacy. Insurance companies, for example, can misuse a patient's medical history, refusing to provide coverage [26]. To address this, some Blockchain and IPFS-based solutions, such as Misbhauddin et al. [27] and Zyskind et al. [28], store sensitive user data off-chain (in IPFS), minimizing personal information exposure risk.

Still, these solutions struggle with medical record sharing between patients and medical centers (i.e., medical staff). To address this, we propose an approach based on Blockchain, smart contract, and NFT technologies. Here, personal information and treatment history are stored as NFTs, and medical test result-related information is stored off-chain. NFTs are generated for each test and shared easily with required addresses (e.g., nurses, doctors). Each patient is assigned a unique identifier to differentiate them from others.

Our work thus provides four key contributions: (a) A Blockchain, smart contract, and NFT-based mechanism for sharing test results. (b) A storage model based on the NFT tool. (c) A proof-of-concept implemented based on the proposed model. and (d) Deployment of the proof-of-concept on four platforms that support ERC721 (NFT of ETH) and EVM (for deploying smart contracts written in solidity language) including BNB Smart Chain, Fantom, Polygon, and Celo to determine the most suitable platform for our proposed model.⁴

The structure of this paper unfolds over eight sections. Following the introductory part, we provide background information, offering an overview of contemporary studies addressing similar research issues and a summary of relevant technologies and EVM-compatible blockchain platforms in Section II. In Section III we explore related work. The subsequent two sections delve into our methodology and the practical implementation of our proposed model (refer to Sections IV, V). To attest to the efficacy of our approach, Section VI presents an evaluation conducted under various scenarios, preceding the discussion in Section VII. Finally, in Section VIII, we provide a summary and outline future directions for this research.

II. BACKGROUND

This section provides a detailed background to the technologies central to the decentralized management of medical test results. Specifically, we explore Blockchain, Smart Contract, Non-Fungible Tokens (NFTs), Ethereum, Binance Smart Chain, Polygon, Celo, and Fantom. Due to the limited scope of this paper, we cannot provide the details on each topic. We prefer the reader follow the white paper/external source if they want to detail the corresponding platforms or topics.

A. Blockchain

Blockchain technology forms the backbone of many decentralized systems, including cryptocurrencies like Bitcoin. It employs a distributed ledger, functioning as a shared database spread across multiple nodes in a network. Each block in the chain contains data, and every new block is linked to the preceding block, forming a chain-like structure. It is the Blockchain's immutable and transparent nature that makes it attractive for various applications, including medical data management.

⁴We did not deploy smart contracts on ETH due to the high execution fees of smart contracts.

B. Smart Contract

Smart contracts are self-executing contracts with the terms of the agreement directly written into lines of code. They eliminate the need for a middleman in digital agreements, ensuring trust, transparency, and efficiency. Smart contracts execute automatically upon meeting predefined rules and conditions. They are stored on the blockchain, making them tamper-proof and traceable. These features make smart contracts a valuable tool in healthcare, specifically in managing and securing patient data.

C. Non-Fungible Tokens (NFTs)

NFTs represent a unique digital asset that is verifiably unique, unlike cryptocurrencies such as Bitcoin or Ethereum, where each unit or "coin" is identical to every other coin. This uniqueness and indivisibility make NFTs ideal for representing ownership or proof of authenticity of individual items or assets, such as artwork, real estate, and in our context, unique sets of medical test results.

D. Ethereum

Ethereum is an open-source blockchain platform that supports smart contract functionality. It provides the underlying technology for a multitude of decentralized applications (DApps). Ethereum also introduced the concept of programmable transactions using smart contracts, making it a pioneer platform for building complex decentralized applications, including those for decentralized healthcare systems.

E. Binance Smart Chain (BSC)

The Binance Smart Chain⁵, an innovation by the Binance cryptocurrency exchange, is a blockchain platform built for running smart contract-based applications. It allows developers to build decentralized applications efficiently and is fully compatible with Ethereum Virtual Machine (EVM). BSC also boasts high transaction speed and low fees, making it a favored platform for various decentralized projects.

F. Polygon

Polygon or MATIC⁶, previously known as Matic Network, is a layer-2 scaling solution for Ethereum. It aims to provide faster and cheaper transactions on the Ethereum blockchain while maintaining its robust security. Polygon uses a technology known as 'sidechains,' which are blockchain systems that run alongside the Ethereum mainchain. This feature allows for scalability, making it suitable for a decentralized medical test result management system.

G. Celo

Celo⁷ is an open blockchain platform that makes financial tools accessible to anyone with a mobile phone. It's designed to support stablecoins and tokenized assets, prioritizing scalability, and usability. Celo's lightweight identity and proof-of-stake mechanisms make it an attractive platform for projects needing secure, fast, and low-cost operations, such as those required in managing medical test results.

⁵<https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md>

⁶<https://polygon.technology/lightpaper-polygon.pdf>

⁷<https://celo.org/papers/whitepaper>

H. Fantom

Fantom⁸ is a high-performance, scalable, and secure smart-contract platform. It is designed to overcome the limitations of previous generation blockchain platforms. Fantom is permissionless, decentralized, and open-source. Its aBFT consensus protocol delivers unparalleled speed, security, and reliability. Fantom's technology stands out for its speed, low transaction costs, and high security, making it a good option for any decentralized application like managing medical test results.

III. RELATED WORK

This section critically surveys the past methodologies employed in creating models for patient test results management, particularly those harnessing Blockchain technology and smart contracts. The study is bifurcated into two core perspectives: i) patient-centric health information management models, and ii) strategies based on blockchain technology.

A. Patient-Oriented Health Information Management Models

In the rapidly evolving healthcare landscape, the patient-centric model has emerged as a pioneering approach, prioritizing patients' needs and values. The key aspect of these models is to cater to patients' privacy preferences, providing them with the ability to have greater control over their health data. The type of data that falls under this model not only includes clinical details essential for disease management and treatment like heart rate, blood pressure, and other vital health indicators, but also personal data like location, phone number, and more. While not all data are directly relevant for treatment, they are critical components of comprehensive patient care.

Among the remarkable contributions in this space is a model introduced by Chen et al. [29]. This innovative approach utilizes Internet of Things (IoT) devices and sensor technology, which are embedded directly into patients. The devices are leveraged to extract vital medical information in real-time. In this model, blockchain technology plays a pivotal role in securely storing, managing, and controlling the harvested data from these IoT devices. The collected information is encrypted before leaving the patient's control and is sent directly to cloud servers.

An intriguing shift away from conventional models is observed in these novel approaches. Traditional models entrust medical centers or hospitals with the storage and management of patient information. In contrast, the newer methodologies propose a paradigm shift, advocating for the patients' power to control their data [30], [31], [32]. This is essentially empowering patients to decide who they want to share their data with, ensuring that data sharing occurs only with trusted entities.

Further substantiating the value of patient empowerment, Makubalo et al. [33] collated several models that endorse health information sharing by patients themselves. In order to prevent the illicit sharing of information by those it has been shared with (like doctors or nurses), a robust system was introduced by Yin et al. [34]. This system employs attribute-based encryption (ABE) to secure data privacy, providing patients the ability to define their data access policies.

This shift towards patient-centered models has fostered a diverse body of work in the area of health data management and privacy. Several other studies have adopted ABE-based access control models, and dynamic policy models to enhance flexibility in the healthcare environment [35], [36], [24]. All these advances reflect the broader move toward empowering patients and improving the flexibility and privacy of health data management.

B. Blockchain-based Health Information Management Models

A distinct, parallel body of research focuses on the implementation of blockchain technology in healthcare data management. In these models, the emphasis is on i) the development of a decentralized management system for patient medical data, inclusive of laboratory information, and ii) the utilization of the InterPlanetary File System (IPFS) for reducing the volume of on-chain stored information.

Madine et al. [37], for instance, proposed a model that stored medical records on a blockchain, preserving the detailed information on IPFS. This approach seeks to achieve a delicate balance between information accessibility and privacy, with a clear objective of safeguarding patient data against unauthorized access from within the same system.

In a similar vein, HealthBank and HealthNautica introduced blockchain and IPFS amalgamations to propose patient-centric models complying with privacy regulations, like the General Data Protection Regulation (GDPR). These models promote the concept of decentralization in healthcare, eliminating the reliance on a central authority for data storage and management, and significantly enhancing data security.

Noteworthy is the evolution of these models beyond simple storage to include sharing of essential information with authorized individuals, such as medical staff at healthcare centers [38], [39]. This represents a more comprehensive approach to health data management, incorporating the need for data sharing in addition to secure storage.

Another pertinent aspect addressed by some studies is the consideration of indirect participants in the treatment process, such as insurance companies and regulators [40]. This approach is especially important as it encompasses the complete ecosystem of healthcare, from patient care to insurance processes and regulatory compliance.

Despite these advances, the field faces a host of challenges. For example, the user-centric model often results in policy redundancy, and the introduction of new blockchain and IPFS-based systems can be complex for users unfamiliar with such technologies.

To circumvent these challenges, our proposed model incorporates a fusion of modern platforms, including blockchain, smart contracts, and Non-Fungible Tokens (NFTs). Rather than depending on sharing policies to define access to test results or patient history, we propose the creation of corresponding NFTs. This approach is designed to alleviate stringent platform requirements, such as those associated with security policy-based methods. The following section elucidates our proposed model that employs NFT technology (i.e., ERC721) to share information with the appropriate entities.

⁸<https://whitepaper.io/document/438/fantom-whitepaper>

IV. BLOCKCHAIN-FACILITATED MEDICAL TEST RESULTS MANAGEMENT FRAMEWORK

In the ensuing discussions, we will initially revisit the traditional methodologies employed in medical test results management. Following this, we introduce our novel model, which is underpinned by Blockchain technology, smart contracts, and Non-Fungible Tokens (NFTs).

A. Conventional Model for Managing Medical Test Results

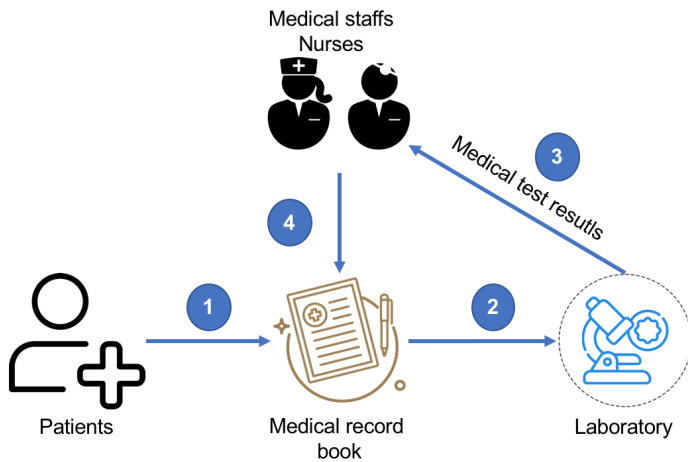


Fig. 1. Conventional medical test results management framework.

Fig. 1 illustrates the conventional procedure for testing and obtaining results, typically structured around four key stages. A patient, in the initial stage, constructs a medical record book, either in electronic or physical form, at the medical institution. This repository houses comprehensive information pertinent to the patient's treatment process and medical test results.

These records are absolutely critical, as the clinicians' diagnostic decisions and medical judgments hinge on the variations observed in a patient's health status as interpreted from these test results. In many developing nations, the testing phase could be quite protracted due to constraints related to infrastructure and auxiliary medical equipment. The patient, consequently, has to endure substantial waiting periods to provide samples and receive the corresponding test results.

Upon receipt of the results from the laboratory personnel, the patient presents these results to the medical practitioners for an assessment of their health status and determining the appropriate therapeutic interventions. All data concerning the diagnosis and resultant treatment are updated in the medical record book.

The loss of a patient's medical record book, thus, severely jeopardizes the treatment process. For the digital variants, the medical record book is stored locally at a discrete hospital or healthcare facility. Given the highly sensitive nature of medical data, the prospects of sharing this information with other institutions are typically slim. Therefore, there exists a pressing need for a comprehensive solution to issues related to the storage and sharing of patients' medical record books, catering to both electronic and physical formats. In the subsequent segment, we unveil our resolution, leveraging the capabilities

of contemporary technologies like blockchain, smart contracts, and NFTs.

B. Medical Test Results Management Model Leveraging Blockchain Technology, Smart Contracts, and NFTs

Fig. 2 exhibits our solution, drawing upon blockchain technology, smart contracts, and NFTs, and comprising of nine critical stages. Users are provided with the ability to create an identifier, referred to as 'patient_ID_global,' which is valid across all medical systems (step 1). This identifier is further linked to a medical record book that archives all relevant details about the medical record, test results, and the patient's medical history (step 2). Steps 3-6 are intrinsically connected with User-Interface (UI) services, offering interfaces to every user group within the system to curtail complex operations (i.e., backend processing). These interactions are facilitated via smart contracts housing pertinent functions for data storage and processing (step 7). In this phase, we devise functions relating to contract creation/NFT or the transfer of NFTs (refer to the introduction).

All transactions are subsequently updated, stored, and dispersed within a distributed ledger, including details about visitors, time, and location, etc. The information relevant to the test results is produced in the corresponding NFTs and transferred to the physicians responsible for treating the patient (step 9).

V. IMPLEMENTATION

The practical application of our model zeroes in on two primary objectives: i) manipulation of data, specifically medical test results, involving creation, query, and update on the blockchain platform, and ii) construction of Non-Fungible Tokens (NFTs) for medical test results, enabling the easy sharing of such data by patients with medical practitioners such as doctors and nurses.

A. Data Creation Procedure

Fig. 3 provides a graphic representation of the steps involved in data initialization with respect to the medical test results. These results incorporate information such as the type of test conducted, time of testing, testing facility, test results, consultation outcomes, and the corresponding treatment approach and its duration. Additionally, metadata of the test results also includes information about the type of patient and the medical personnel involved in conducting the test.

The storage process, in this context, facilitates concurrent storage (i.e., distributed processing as a peer-to-peer network) on a distributed ledger, which supports multiple users for concurrent storage, thereby reducing system latency.

In essence, the medical test results data is structured as follows:

```
medicalTestResultsObject = {  
  "patientID": patientID,  
  "medicalTestID": medicalTestID,  
  "medicalStaffID": medicalStaffID,  
  "type": type of test,  
  "numbers": numbers of treatments,  
}
```

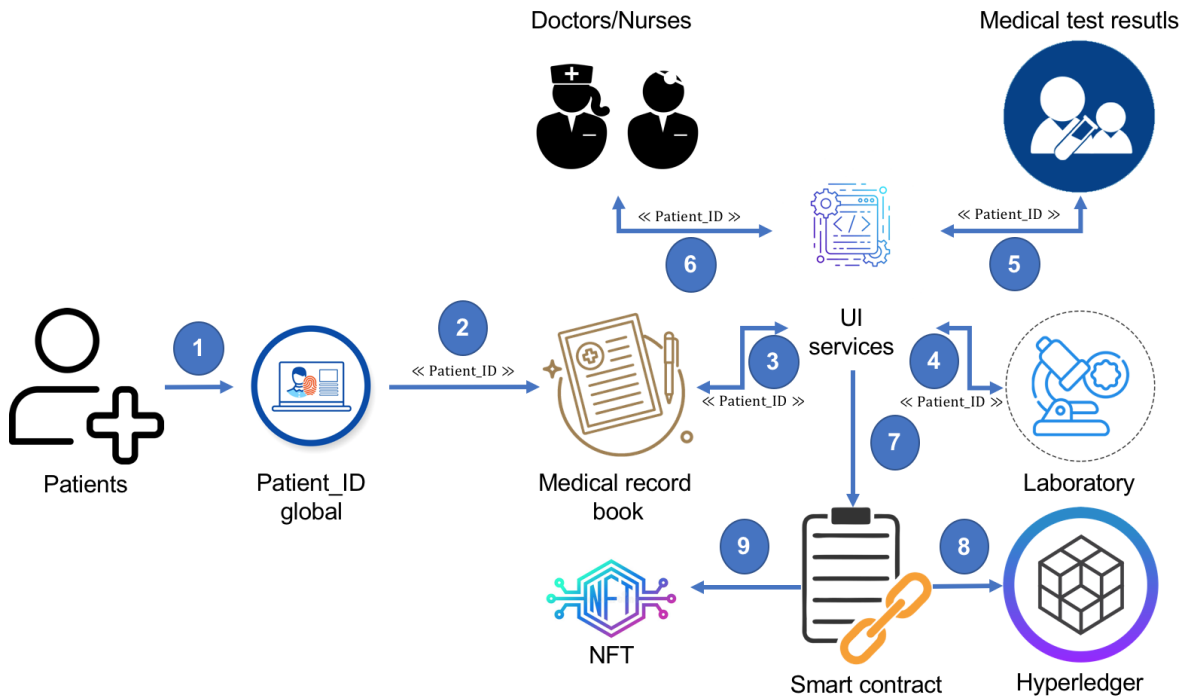


Fig. 2. Medical test results management model leveraging blockchain technology, smart contracts, and NFTs.

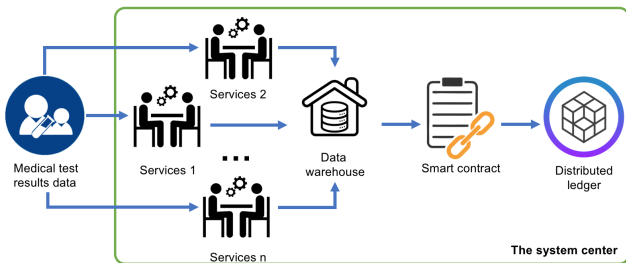


Fig. 3. Initialization of data and Non-Fungible Tokens (NFTs).

```

"results": results of the medical test,
"diagnose": diagnosis of the illness,
"institution": institutionID,
"date": time and date,
"times": times of test,
"period": period of the treatment,
"state": Null
};

```

In particular, besides information useful for content extraction (i.e., medical staff, test results, diagnostic outcomes, etc.), we also store information related to the status of the patient's treatment at the hospital (i.e., "state" - default set to Null). Specifically, the "state" changes to a value of 1 when the respective patient has completed their treatment and exited the medical facility (i.e., numbers increment by 1); a value of 0 signifies that the patient is still under treatment. Furthermore, we keep a record of the treatment interval and number of tests conducted through two parameters: "period" and "times".

Following this, pre-designed constraints in the Smart Con-

tract are invoked through the API (i.e., name of function) to synchronize them up the chain. This role of validation carries significant weight as it directly impacts the process of storing medical information (i.e., medical test results), as well as the treatment of patients.

For processes that initialize NFTs (i.e., store only test results), the contents of the NFT are defined as follows:

```

NFT MEDICAL_RECORD = {
"medicalRecordID": medicalRecordID,
"patientID": patientID,
"medicalTestID": medicalTestID,
"type": type of test,
"medicalStaffID": medicalStaffID,
"results": results of the medical test,
"institution": institutionID,
"date": time and date,
};

```

The above-mentioned information is extracted from the original data stored on the chain - our previous model constructed a role-based access control (RBAC) system, hence, direct access from non-owners or unauthorized entities is not possible. Also, considering that a patient undergoes several health assessments/checks before a disease is diagnosed, the information extracted minimizes the risk of data loss. For instance, a doctor diagnosing blood issues does not need access to a patient's bone X-ray.

B. Data Access Procedure

Mirroring the process of data initialization, the method of data retrieval also allows multiple participants to concurrently access the system (i.e., in a distributed model). Assistance

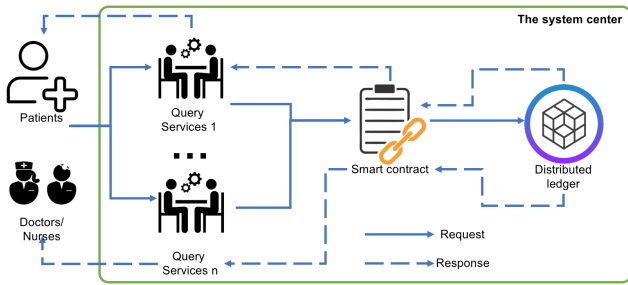


Fig. 4. Data access procedure.

services manage requests coming from medical professionals (like nurses and doctors) and patients who seek to view the data. Depending on the identity of the person making the request, the objectives of access may vary. Specifically, medical personnel might query to validate the procedure of medical testing (i.e., test outcomes), whereas patients might wish to seek details about the current holders of their NFTs.

Fig. 4 showcases the stages involved in retrieving medical test outcome data. These requests are conveyed as services (i.e., pre-configured APIs) from the requester to the existing smart contracts in the system (i.e., function names) before fetching the data from the distributed ledger. All retrieval demands are also kept as access history for each person or entity involved.

If the relevant information cannot be traced (e.g., incorrect ID), the system will return a “results not found” message. Regarding NFT access procedures, all assistance services are provided in the form of APIs.

C. Data Update Procedure

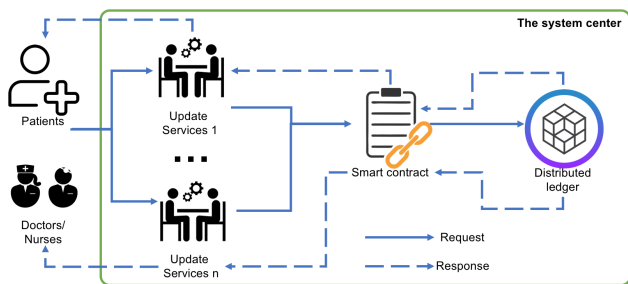


Fig. 5. Data update procedure.

The process of data modification commences only after verifying the existence of the data on the blockchain (i.e., post the corresponding data access procedure). In this segment, we assume that the searched data is present on the blockchain. If the data does not exist, the system sends a “results not found” message to the user (see V-B for further details).

Like the procedures of data access and initialization, we offer modification services as APIs to receive user requests before forwarding them to the smart contract (i.e., function name) for execution. This procedure aims to update test results to minimize patient waiting periods in healthcare institutions. Moreover, it aids doctors in tracing their treatment path based on the associated sequence of NFTs.

Fig. 5 demonstrates the process for modifying medical test results. Concerning NFTs (i.e., available), the update process involves only the transfer from the current holder’s address to a new one (i.e., new holder). If any update is made to an existing NFT, it will be registered as a new NFT (refer to V-A for further details).

VI. EVALUATION SCENARIOS

The model for generating and managing medical test results is designed to simplify the process for patients. It allows easy management and sharing of medical records with relevant parties. Rather than solely relying on traditional security policies such as access control, we harness the robust and transparent nature of blockchain technology. We chose to leverage Ethereum Virtual Machine (EVM)-enabled blockchain platforms over the Hyperledger ecosystem for its wider accessibility and utilization in existing platforms and systems. Previously, we had assessed system responsiveness, including the number of requests responded to successfully or failed and system latency.

In this paper, we delve into determining the most suitable platform for our proposed model based on economic considerations. Specifically, we implemented a prototype system on four renowned blockchain platforms that support Ethereum Virtual Machine (EVM). The selected platforms include Binance Smart Chain (BNB Smart Chain), Polygon, Fantom, and Celo. Not only did we analyze the performance and cost-effectiveness of these platforms, but we also shared our implementation as a contribution to the wider community.

In these implementations, transaction fees correspond to the supporting coins of the respective platforms. The models were implemented on the 24th of November, 2022 at 8:44:53 AM UTC, with the fees paid in BNB, MATIC, FTM, and CELO, respectively.

A. Binance Smart Chain Implementation and Analysis (sample)

Transaction Hash	Method	Block	Age	From	To	Value	[Txn Fee]
0x04205158931273143d4...	Transfer	24867369	1 day 17 hrs ago	0xc0a8c5d44206e0834f...	0xaf33888d1dfb6957b1...	0 BNB	0.00057003
0x1f5ae508ae1c00322...	Mint	24867361	1 day 18 hrs ago	0xc0a8c5d44206e0834f...	0xaf33888d1dfb6957b1...	0 BNB	0.00101842
0xb0c03161087984cc8...	Contract Creation	24867375	1 day 18 hrs ago	0xc0a8c5d44206e0834f...		0 BNB	0.027134

Fig. 6. The transaction info for BNB smart chain.

We initiated our assessment by implementing the model on the Binance Smart Chain (BNB Smart Chain), and a detailed exploration of the successful installation on this chain is presented in Fig. 6. Binance Smart Chain, as an offshoot of the Binance Chain, offers EVM-compatibility, allowing us to deploy Solidity-based smart contracts with relative ease. It further benefits from its parent chain’s high-speed transactions while allowing for better decentralization.

B. NFT Creation Process

Following successful implementation, we explored NFT generation, a pivotal aspect of the proposed model. Fig. 7

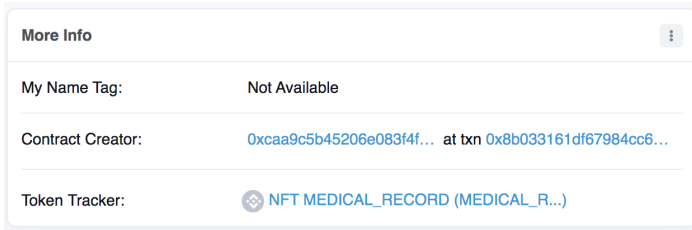


Fig. 7. NFT creation on binance smart chain.

visually demonstrates the creation of an NFT on the Binance Smart Chain. NFTs, inherently unique in their characteristics, perfectly capture the distinct nature of individual medical test results and patient records.

C. NFT Transfer

In line with NFT creation, the retrieval and transfer of these tokens represent the next step in the model's operation. As depicted in Fig. 8, the transfer of NFT ownership addresses occurs smoothly on the Binance Smart Chain. This transferability of NFTs is essential for enabling patients to share their medical test results or records with medical professionals or any other entities of their choice.

In the following sections, we delve into evaluating the transactional aspects of the model, focusing on factors such as transaction fee, gas limit, gas used by the transaction, and gas price. To obtain a comprehensive view of the performance and cost implications of the proposed model, we replicated the same set of operations on the other three selected platforms, namely Polygon, Fantom, and Celo. The underlying motive is to identify the most cost-effective platform for deployment.

Similar settings were used in the remaining platforms to allow for a fair comparison, and the subsequent assessments offer a detailed examination of the transactional metrics. For more detail we refer the reader to check our deployment of the smart contract on the four EVM-supported platforms, namely BNB⁹; MATIC¹⁰; FTM¹¹; and CELO¹².

D. Transaction Fee

In Table I, we dissect the transaction costs associated with creating contracts on all four platforms. It is clearly evident that the most capital-intensive operation across the platforms is contract creation, with BNB Smart Chain exhibiting the highest cost of 0.0273134 BNB (\$8.43). Conversely, Fantom's platform reported the most economical contract initiation fee, standing at less than 0.00957754 FTM (\$0.001849). The transaction fee for contract creation on Celo's platform was marginally cheaper than that of Polygon, totaling only \$0.004 compared to Polygon's \$0.01.

⁹<https://testnet.bscscan.com/address/0xafa3888d1dfbf957b1cd68c36\ede4991e104a53>

¹⁰<https://mumbai.polygonscan.com/address/0xd9ee80d850ef3c4978dd0b099a45a559fd7c5ef4>

¹¹<https://testnet.ftmscan.com/address/0x4a2573478c67a894e32d\806c8dd23ee8e26f7847>

¹²<https://explorer.celo.org/alfajores/address/0x4a2573478C67a894E32D806c8Dd23EE8E26f7847/transactions>

Turning our attention to the subsequent two operations, Create NFT and Transfer NFT, the associated costs on all three platforms (Polygon, Celo, and Fantom) are remarkably low, verging on negligible. In stark contrast, the transaction cost on BNB Smart Chain remains considerably higher, amounting to 0.00109162 BNB (\$0.34) and 0.00057003 BNB (\$0.18) for Create NFT and Transfer NFT, respectively. This disparity underscores the need for an in-depth economic evaluation when selecting a suitable platform for blockchain-based solutions.

E. Gas Limit

Table II showcases the gas limit for each transaction across the platforms. The gas limits for BNB, Polygon, and Fantom remain relatively equivalent, with Polygon and Fantom displaying almost identical figures across all transaction types. Celo, however, sets a significantly higher gas limit, amounting to 3,548,922; 142,040; and 85,673 for contract creation, NFT creation, and NFT transfer, respectively. This discrepancy can have notable implications on the transactional performance and cost-effectiveness of deploying solutions on these platforms.

F. Gas Used by Transaction

Table III illustrates the proportion of the total gas limit consumed by each transaction, as per the figures displayed in Table II. It is noteworthy that BNB, Polygon, and Fantom utilized 100% of the allocated gas limit for the operations of contract creation and NFT creation. Celo's utilization, on the other hand, amounted to 76.92% of the gas limit for these two transactions. When observing the NFT transfer transaction, Fantom and Polygon recorded the highest gas consumption levels at 93.41% of the gas limit, whereas BNB and Celo's consumption stood at 79.17% and 69.8% respectively.

G. Gas Price

As shown in Table IV, the gas price across the platforms remained relatively stable for each transaction type. The BNB Smart Chain showcased the highest gas price, measuring 10 Gwei for all transaction types. Conversely, the Polygon and Celo platforms reflected the lowest gas prices at 2.500000012 and 2 Gwei respectively. Fantom's platform priced gas at 3.5 Gwei, marginally higher than Polygon and Celo, yet significantly lower than the BNB Smart Chain.

H. Summary

In summation, this comparative analysis underscores the essentiality of in-depth cost evaluation before choosing a platform to deploy blockchain solutions. The BNB Smart Chain emerges as the most expensive platform for all transaction types, while Polygon, Celo, and Fantom provide significantly cheaper alternatives. Particularly, Fantom and Polygon offer remarkably low transaction fees and competitive gas limits, making them potentially viable choices for the cost-effective deployment of NFTs. However, consideration should also be given to other critical factors such as platform maturity, ecosystem support, developer experience, security, and user adoption, which are outside the scope of this analysis.

Txn Hash	Age	From	To	Token ID	Token
0x0d2615893127314da4...	1 day 18 hrs ago	0xafaa3888d1dfbfe957b1...	OUT 0xcaa9c5b45206e083f4f...	1	ERC-721: NFT.....ORD
0x1fb5ae508ae1c00322...	1 day 18 hrs ago	0x00000000000000000000...	IN 0xafaa3888d1dfbfe957b1...	1	ERC-721: NFT.....ORD

[\[Download CSV Export\]](#)

Fig. 8. NFT transfer on binance smart chain.

TABLE I. TRANSACTION FEE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.0273134 BNB (\$8.43)	0.00109162 BNB (\$0.34)	0.00057003 BNB (\$0.18)
Fantom	0.00957754 FTM (\$0.001849)	0.000405167 FTM (\$0.000078)	0.0002380105 FTM (\$0.000046)
Polygon	0.006840710032835408 MATIC(\$0.01)	0.000289405001852192 MATIC(\$0.00)	0.000170007501088048 MATIC(\$0.00)
Celo	0.007097844 CELO (\$0.004)	0.0002840812 CELO (\$0.000)	0.0001554878 CELO (\$0.000)

TABLE II. GAS LIMIT

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,340	109,162	72,003
Fantom	2,736,440	115,762	72,803
Polygon	2,736,284	115,762	72,803
Celo	3,548,922	142,040	85,673

TABLE III. GAS USED BY TRANSACTION

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,340 (100%)	109,162 (100%)	57,003 (79.17%)
Fantom	2,736,440 (100%)	115,762 (100%)	68,003 (93.41%)
Polygon	2,736,284 (100%)	115,762 (100%)	68,003 (93.41%)
Celo	2,729,940 (76.92%)	109,262 (76.92%)	59,803 (69.8%)

VII. DISCUSSION

A. Notable Observations

Our research involving the evaluation and comparison of transaction costs, gas limits, and gas prices across four well-known EVM-compatible blockchain platforms - Binance Smart Chain (BNB), Fantom, Polygon, and Celo - has surfaced several noteworthy findings.

Foremost among them is the cost implication associated with different blockchain platforms. We observed that the Binance Smart Chain tended to levy the highest transaction fees and gas prices, thus potentially raising the cost of blockchain operations for developers and users on this platform. In sharp contrast, Fantom, Polygon, and Celo proved to be more cost-friendly alternatives, with Fantom presenting the lowest transaction fees among the four platforms.

Interestingly, despite lower fees, the gas limits on Polygon, Fantom, and Celo were not vastly different from that of Binance Smart Chain, suggesting that these platforms could potentially be matching the service levels of Binance Smart Chain while also being more economically efficient.

This underscores an important trade-off that users, developers, and companies need to consider when choosing between these platforms: While Binance Smart Chain might be more established and widely accepted, newer platforms like Fantom, Polygon, and Celo are providing compelling value propositions

in terms of cost efficiencies, which could lead to considerable savings in the long run.

B. Threats to Validity

Despite the informative nature of our study, it is crucial to recognize the limitations and potential threats to its validity.

1) *Temporal volatility*: The world of blockchain and cryptocurrencies is notably volatile, and costs associated with transactions, gas limits, and gas prices are dynamic, changing in response to a multitude of factors such as market conditions, supply and demand dynamics, among others. Hence, the values presented in this paper may change over time, and users are advised to consider the most recent data while making decisions.

2) *Network variations*: The state of the network at the time of evaluation could significantly impact the results. Network congestion, often arising due to a surge in demand for transactions, typically leads to an increase in fees, and the reverse is true during periods of lower demand.

3) *Platform-Specific variables*: Each blockchain platform is uniquely designed, having its own set of characteristics including consensus mechanisms, block time, network size, and more. All these factors can greatly influence transaction costs and gas limits, and our study does not account for these platform-specific variables.

4) *Scope constraints*: Our analysis included only a limited number of EVM-compatible blockchain platforms and transaction types. Including additional platforms and a wider variety of transaction types could potentially yield different insights and conclusions.

C. Directions for Future Research

The findings of this study open up a multitude of interesting directions for future research:

1) *Real-time cost analysis*: Given the rapid changes in transaction costs in the realm of blockchain, future research could develop a real-time or dynamic analysis model that captures the cost fluctuations across different platforms over a defined period. This could provide more current and actionable insights for users and developers.

TABLE IV. GAS PRICE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)
Fantom	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)
Polygon	0.000000002500000012 MATIC (2.500000012 Gwei)	0.000000002500000016 MATIC (2.500000016 Gwei)	0.000000002500000016 MATIC (2.500000016 Gwei)
Celo	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)

2) *Comprehensive performance evaluation:* While our study focused primarily on costs, further research could explore other performance metrics such as transaction speed, scalability, security, and reliability. A comprehensive evaluation using multiple performance indicators could help users make a more informed choice of blockchain platform based on their specific needs.

3) *Inclusion of other blockchain platforms:* Our analysis was confined to a selected few EVM-compatible blockchains. Future research can incorporate more blockchain platforms, broadening the scope of comparison, and providing a more diverse range of options for users to consider.

4) *Application-specific evaluation:* Another interesting direction could be to investigate the cost and feasibility of deploying specific applications, such as decentralized finance (DeFi), supply chain management, gaming, and more across different blockchain platforms. An application-specific analysis could provide more targeted insights for developers and stakeholders in these domains.

5) *Privacy and efficiency implications:* In our present examination, we have yet to explore issues associated with the privacy policy of users, such as access control [26], [36] or dynamic policy [41], [42]. These aspects represent potential pathways for future research endeavors. Lastly, methodologies grounded in infrastructure (such as gRPC [43], [44]; Microservices [45], [46]; Dynamic message transmission [47] and Brokerless mechanisms [48]) could be incorporated into our model to boost user interaction, specifically through an API-call-based approach.

In general, our study provides a valuable comparative analysis of transaction costs across different blockchain platforms, which can serve as a useful resource for developers, businesses, and researchers alike. As this area continues to evolve at a rapid pace, continuous monitoring and analysis are crucial to keep up-to-date with the latest developments. Our study provides a foundation upon which more comprehensive, real-time, and application-specific analyses can be built in the future.

VIII. CONCLUSION

In this paper, we have analyzed and compared transaction costs across four prominent EVM-compatible blockchain platforms - Binance Smart Chain (BNB), Fantom, Polygon, and Celo. We evaluated the costs from multiple perspectives, including transaction fees, gas limits, gas used by transaction, and gas prices.

Our research findings reveal significant differences in the transaction cost structure across these platforms. Binance Smart Chain surfaced as the most expensive, with the highest

transaction fees and gas prices, while Fantom offered the lowest transaction costs. However, the gas limits across the platforms were comparable, signifying that less expensive platforms could provide a similar level of service as Binance Smart Chain, but at a lower cost. While our research provides valuable insights, it is subject to several limitations, primarily due to the dynamic and rapidly evolving nature of the blockchain landscape. The cost parameters we evaluated are subject to market fluctuations, network congestion levels, and platform-specific variables. Therefore, the results should be interpreted with caution, and users are advised to consider the most current data while making decisions.

Our study opens the door for further research in this domain. Future work could include real-time cost analysis, a comprehensive evaluation of multiple performance metrics, inclusion of more blockchain platforms, and application-specific evaluations. In conclusion, our research underscores the importance of considering transaction costs while choosing a blockchain platform. It provides a clear direction for developers, companies, and researchers, helping them make informed decisions that balance cost and performance. As the blockchain ecosystem continues to grow and evolve, studies like ours will become increasingly crucial in navigating the complex landscape of blockchain platforms and their associated cost structures.

REFERENCES

- [1] C.-J. Su and C.-Y. Wu, "Jade implemented mobile multi-agent based, distributed information platform for pervasive health care monitoring," *Applied Soft Computing*, vol. 11, no. 1, pp. 315–325, 2011.
- [2] M. Kyrarini, F. Lygerakis, A. Rajavenkatanarayanan, C. Sevastopoulos, H. R. Nambiappan, K. K. Chaitanya, A. R. Babu, J. Mathew, and F. Makedon, "A survey of robots in healthcare," *Technologies*, vol. 9, no. 1, p. 8, 2021.
- [3] W. H. Organization, *Oral health surveys: basic methods*. World Health Organization, 2013.
- [4] A. Rais and A. Viana, "Operations research in healthcare: a survey," *International transactions in operational research*, vol. 18, no. 1, pp. 1–31, 2011.
- [5] K. S. Chan, J. B. Fowles, and J. P. Weiner, "Electronic health records and the reliability and validity of quality measures: a review of the literature," *Medical Care Research and Review*, vol. 67, no. 5, pp. 503–527, 2010.
- [6] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, "Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 71–78.
- [7] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.
- [8] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.

- [9] —, “Bitcoin: A peer-to-peer electronic cash system bitcoin: A peer-to-peer electronic cash system,” *Bitcoin.org. Disponible en https://bitcoin.org/en/bitcoin-paper*, 2009.
- [10] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, “Multi-sessions mechanism for decentralized cash on delivery system,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019.
- [11] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, “Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems,” in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020, 2020, p. 105–109.
- [12] —, “On components of a patient-centered healthcare system using smart contract,” in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, 2020, p. 31–35.
- [13] H. T. Le, K. L. Quoc, T. A. Nguyen, K. T. Dang, H. K. Vo, H. H. Luong, H. Le Van, K. H. Gia, L. V. Cao Phu, D. Nguyen Truong Quoc *et al.*, “Medical-waste chain: A medical waste collection, classification and treatment management by blockchain technology,” *Computers*, vol. 11, no. 7, p. 113, 2022.
- [14] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, “Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies,” in *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer, 2022, pp. 576–583.
- [15] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, “Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems,” in *International Conference on Mobile, Secure, and Programmable Networking*. Springer, 2020, pp. 44–56.
- [16] N. H. Tuan Khoi *et al.*, “Vblock - blockchain based traceability in medical products supply chain management: Case study in vietnam,” in *International Conference on Artificial Intelligence for Smart Community*, 2020.
- [17] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, “Toward a design of blood donation management by blockchain technologies,” in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 78–90.
- [18] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, “Blockchain: A blood donation network managed by blockchain technologies,” *Network*, vol. 2, no. 1, pp. 21–35, 2022.
- [19] H. X. Son, M. H. Nguyen, N. N. Phien, H. T. Le, Q. N. Nguyen, V. Dinh, P. Tru, and P. Nguyen, “Towards a mechanism for protecting seller’s interest of cash on delivery by using smart contract in hyperledger,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 4, pp. 45–50, 2019.
- [20] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, “Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.
- [21] H. H. Luong, T. K. N. Huynh, A. T. Dao, and H. T. Nguyen, “An approach for project management system based on blockchain,” in *International Conference on Future Data and Security Engineering*. Springer, 2021, pp. 310–326.
- [22] N. H. Tuan Khoi *et al.*, “Domain name system resolution system with hyperledger fabric blockchain,” in *International Conference on Inventive Computation and Information Technologies*, 2022.
- [23] H. T. Le, N. T. T. Le, N. N. Phien, and N. Duong-Trung, “Introducing multi shippers mechanism for decentralized cash on delivery system,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.
- [24] Q. N. T. Thi, T. K. Dang, H. L. Van, and H. X. Son, “Using json to specify privacy preserving-enabled attribute-based access control policies,” in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 561–570.
- [25] J. Abou Jaoude and R. G. Saade, “Blockchain applications—usage in different domains,” *IEEE Access*, vol. 7, pp. 45 360–45 381, 2019.
- [26] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, “Toward an privacy protection based on access control model in hybrid cloud for healthcare systems,” in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on European Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.
- [27] M. Misbhauddin, A. AlAbdulatheam, M. Aloufi, H. Al-Hajji, and A. Al-Ghuwainem, “Medaccess: A scalable architecture for blockchain-based health record management,” in *2020 2nd International Conference on Computer and Information Sciences (ICIS)*. IEEE, 2020, pp. 1–5.
- [28] G. Zyskind, O. Nathan *et al.*, “Decentralizing privacy: Using blockchain to protect personal data,” in *2015 IEEE Security and Privacy Workshops*. IEEE, 2015, pp. 180–184.
- [29] Z. Chen, W. Xu, B. Wang, and H. Yu, “A blockchain-based preserving and sharing system for medical data privacy,” *Future Generation Computer Systems*, vol. 124, pp. 338–350, 2021.
- [30] M. Du, Q. Chen, J. Xiao, H. Yang, and X. Ma, “Supply chain finance innovation using blockchain,” *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1045–1058, 2020.
- [31] M. R. Patra, R. K. Das, and R. P. Padhy, “Crhis: cloud based rural healthcare information system,” in *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance*, 2012, pp. 402–405.
- [32] C. O. Rolim, F. L. Koch, C. B. Westphall, J. Werner, A. Fracalossi, and G. S. Salvador, “A cloud computing solution for patient’s data collection in health care institutions,” in *2010 Second International Conference on eHealth, Telemedicine, and Social Medicine*. IEEE, 2010, pp. 95–99.
- [33] T. Makubalo, B. Scholtz, and T. O. Tokosi, “Blockchain technology for empowering patient-centred healthcare: A pilot study,” in *Conference on e-Business, e-Services and e-Society*. Springer, 2020, pp. 15–26.
- [34] Y. Zhang, M. Qiu, C.-W. Tsai, M. M. Hassan, and A. Alamri, “Healthcps: Healthcare cyber-physical system assisted by cloud and big data,” *IEEE Systems Journal*, vol. 11, no. 1, pp. 88–95, 2015.
- [35] N. M. Hoang and H. X. Son, “A dynamic solution for fine-grained policy conflict resolution,” in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 116–120.
- [36] H. X. Son and N. M. Hoang, “A novel attribute-based access control system for fine-grained privacy protection,” in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.
- [37] M. M. Madine, A. A. Battah, I. Yaqoob, K. Salah, R. Jayaraman, Y. Al-Hammadi, S. Pesic, and S. Ellahham, “Blockchain for giving patients control over their medical records,” *IEEE Access*, vol. 8, pp. 193 102–193 115, 2020.
- [38] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, “Fhirchain: applying blockchain to securely and scalably share clinical data,” *Computational and structural biotechnology journal*, vol. 16, pp. 267–278, 2018.
- [39] V. Patel, “A framework for secure and decentralized sharing of medical imaging data via blockchain consensus,” *Health informatics journal*, vol. 25, no. 4, pp. 1398–1411, 2019.
- [40] M. Kassab, J. DeFranco, T. Malas, P. Laplante, G. Destefanis, and V. V. G. Neto, “Exploring research in blockchain for healthcare and a roadmap for the future,” *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 4, pp. 1835–1852, 2019.
- [41] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, “Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies,” in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.
- [42] H. X. Son, T. K. Dang, and F. Massacci, “Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies,” in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.
- [43] L. T. T. Nguyen *et al.*, “Bmdd: a novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages),” *PeerJ Computer Science*, vol. 8, p. e950, 2022.

- [44] L. N. T. Thanh *et al.*, "Toward a security iot platform with high rate transmission and low energy consumption," in *International Conference on Computational Science and its Applications*. Springer, 2021.
- [45] —, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management*. Springer, 2021.
- [46] L. N. T. Thanh, N. N. Phien, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2021.0120768>
- [47] L. N. T. Thanh *et al.*, "Uip2sop: A unique iot network applying single sign-on and message queue protocol," *IJACSA*, vol. 12, no. 6, 2021.
- [48] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," 2021.

Leveraging Blockchain, Smart Contracts, and NFTs for Streamlining Medical Waste Management: An Examination of the Vietnamese Healthcare Sector

Triet M. N.¹, Khanh H. V.¹, Huong H. L.*¹, Khiem H. G.¹, Phuc T. N.¹, Ngan N. T. K.²,
Quy T. L.¹, Bang L. K.¹, Trong D. P. N.¹, Hieu M. D.¹, Bao Q. T.¹, Khoa D. T.¹, and Anh T. N.¹

¹FPT University, Can Tho City, Viet Nam

²FPT Polytechnic, Can Tho City, Viet Nam

Abstract—Medical waste is deemed hazardous due to its potential health implications and the predominant practice of discarding it post six months of utilization. Furthermore, the reusable proportion of such waste is minimal. The implications of this scenario were brought to the fore during the COVID-19 pandemic when sub-optimal medical waste management was identified as a factor exacerbating the spread of the virus worldwide. The predicament is particularly grave in developing nations, such as Vietnam, where the underdeveloped state of medical infrastructure renders efficient waste management a daunting task. The waste management challenge also stems from the significant roles played by different stakeholders (healthcare workers and patients confined to isolation wards), whose actions directly influence waste classification, impact the waste treatment process, and indirectly contribute to environmental pollution. Given that waste management involves a chain of activities requiring the coordinated efforts of medical, transportation, and waste treatment personnel, inaccuracies in the initial stages, such as waste sorting, can negatively impact subsequent processes. In light of these issues, our study puts forth a unique model aimed at enhancing waste classification and management practices in Vietnam. This model innovatively integrates Blockchain technology, smart contracts, and non-fungible tokens (NFTs) with the intent to foster an increased individual and collective consciousness towards effective waste classification within healthcare settings. Our research is notable for its four-fold contribution: (a) suggesting a unique mechanism based on blockchain technology and smart contracts, designed specifically to improve medical waste classification and treatment in Vietnam; (b) introducing a model for instituting rewards or penalties based on NFT technology to influence behaviors of individuals and organizations; (c) demonstrating the feasibility of the proposed model through a proof-of-concept; (d) executing the proof-of-concept on four prominent platforms that support ERC721 - NFT of Ethereum and EVM for executing smart contracts programmed in the Solidity language, namely BNB Smart Chain, Fantom, Polygon, and Celo.

Keywords—Medical waste management; blockchain; smart contracts; NFTs; ethereum; fantom; polygon; binance smart chain

I. INTRODUCTION

The threat posed by medical waste, a hazardous byproduct of healthcare activities, is of global concern. The vast majority of medical supplies and equipment – nearly 99% – become waste within six months of initial use due to their potential for transmitting infections [1], [2]. The environmental hazards posed by single-use items such as medical gloves, protective gear, and masks, further exacerbate the issue [3]. As such,

regulatory bodies worldwide have implemented stringent processes to ensure proper classification and treatment of medical waste.

A notable facet of this global waste management challenge is the intersection of environmental and economic implications [4]. Materials difficult to break down contribute to pollution, applying immense pressure on the environment [5]. This concern is particularly acute in developing nations where waste disposal processes have shown links to environmental pollution, as seen in India [6] and Brazil [7]. The urgency of the issue further intensified during the pandemic, with the surge in medical supplies leading to increased waste [8], [9].

The Vietnam context presents a unique case. Systematic studies have delved into the role of waste segregation in managing the Covid-19 crisis. However, much of the focus remains on the results or consequences of waste management rather than providing an improved, technologically advanced model aimed at enhancing transparency and decentralized data storage.

Addressing this need, recent research has pivoted towards models utilizing Blockchain technology and smart contracts for waste treatment and classification [10], [11]. Such models focus on identifying the origin and composition of waste and encompass key stakeholders – healthcare workers, patients, waste collectors, and waste treatment companies. Information related to these user groups and waste (referred to as ‘bags’) is validated before being recorded on the chain. This method not only helps pinpoint the source of waste but also minimizes contact between parties, thereby reducing disease transmission risks [12]. As such, these models could supersede traditional waste treatment methods, particularly during epidemic periods.

Additionally, the role of public awareness and cooperation in waste management is crucial to curbing treatment times. The process of self-classification, despite being commonplace in developed nations, only emerged in developing countries following the outbreak of the Covid-19 pandemic. In Vietnam, a large portion of waste is unclassified, significantly impacting its treatment process. Consequently, our research seeks to address this issue by proposing a model for managing medical waste using Blockchain technology and smart contracts. Simultaneously, we aim to shape public perceptions of waste classification by leveraging non-fungible token (NFT) technology.

This study focuses on evaluating existing waste treatment models in developing countries, specifically Vietnam, during the Covid-19 pandemic. It seeks to provide a suitable approach for potential future epidemics. Our main contribution lies in presenting an NFT-based (ERC 721) approach and a penalty system for violations of waste classification norms.

Thus, the four-fold contribution of our work includes: (a) proposing a medical waste classification and treatment mechanism for the Vietnamese context, leveraging blockchain technology and smart contracts; (b) introducing a reward/punishment system based on NFT technology aimed at individuals and organizations; (c) implementing a proof-of-concept of the proposed model using smart contracts; and (d) deploying the proof-of-concept on four platforms supporting ERC721 - NFT of Ethereum and EVM for executing smart contracts written in Solidity, namely BNB Smart Chain, Fantom, Polygon, and Celo.¹

This paper is structured into seven subsequent sections. Following this introduction, we offer a brief overview of Blockchain, Smart contract, EVM, NFT, and the four EVM-supported blockchain platforms in Section II. Then we review related work exploring similar research problems in Section III. Next, we describe our proposed approach and its implementation (Sections IV, V). Section VI demonstrates the effectiveness of our model in different scenarios, followed by a discussion in Section VII. Finally, Section VIII summarizes our work and outlines potential avenues for future research.

II. BACKGROUND

A. Blockchain Technology

Originally conceived as the underlying technology for Bitcoin [14], blockchain has gained recognition for its potential beyond cryptocurrency [15], [16]. Blockchain is often characterized as a transparent, reliable, and decentralized ledger that operates on a peer-to-peer network [17], [18]. It manages transaction data across several computers concurrently, fostering a trust environment that permits autonomous interaction without reliance on a centralized authority [19]. Key benefits of blockchain-based systems include:

- Security: Through digital signatures and encryption, blockchain systems ensure data security and integrity [20].
- Fraud control: Data duplication across multiple nodes provides robust defense against hacking, enabling efficient recovery of records[21].
- Transparency: Real-time transaction status visibility fosters reliability and convenience for all parties involved [22].
- No hidden fees: The decentralized nature eliminates the need for intermediaries, thereby reducing associated costs and commissions.
- Access levels: Users can opt for a public blockchain network accessible to all, or a permissioned network, which requires user authorization for each node[23].

¹We exclude ETH from our deployment because of its prohibitively high smart contract execution fee.[13]

- Speed: Blockchain transactions are expedited due to the lack of external payment system integration, leading to cost and time efficiency[10].
- Account reconciliation: Authenticity and validity of participants are collaboratively verified by the network participants.

B. Smart Contract

Smart contracts, or chaincodes[24], [25], are self-executing contracts where terms of agreement between parties are directly written into lines of code and automated via blockchain technology. Noteworthy characteristics of smart contracts include:

- Distributed: Smart contracts are replicated and distributed across all nodes of the blockchain network, fostering decentralization.
- Deterministic: Smart contracts execute actions as designed under defined conditions, and yield consistent results irrespective of the executor.
- Automate: Capable of automating various tasks, smart contracts operate as self-actuating programs that remain idle until activated.
- Non-modifiable: Post-deployment modifications to smart contracts are impossible. Deletion is possible only if this functionality was predefined.
- Customizable: Smart contracts can be programmed diversely before deployment, enabling the creation of various types of decentralized applications (Dapps).
- Trust-less interactions: Smart contracts allow parties to interact without mutual trust, as blockchain technology ensures data accuracy.
- Transparency: As smart contracts operate on a public blockchain, their source code is immutable and publicly viewable.

C. Blockchain Platforms

1) *Ethereum*: Ethereum [26] is a decentralized platform that supports the development and execution of smart contracts via Turing-complete programming languages. These smart contracts are executed by the Ethereum Virtual Machine (EVM) and can be written in languages such as Solidity, Serpent, Low-level Lisp-like Language (LLL), and Mutan. Ethereum enables the creation of various applications, including financial contracts, betting markets, and withdrawal limits. As of now, it remains the most popular platform for smart contract development.

D. Ethereum Virtual Machine (EVM)

The Ethereum Virtual Machine (EVM) is a Turing-complete software that operates as a runtime environment for smart contracts in Ethereum. It is completely isolated from the main Ethereum network, which makes it a perfect sandbox for running untrusted code [27]. As such, smart contracts can't communicate with other contracts directly. Instead, they do so

via the EVM, preventing any potential malicious code from affecting the network.

When a smart contract is executed, each and every instruction is run on every node in the network. This redundancy helps ensure the security and robustness of the network, but it also necessitates a mechanism for restricting resource consumption on the network. To this end, Ethereum implements a system known as “gas” – each instruction requires a certain amount of gas to execute. Gas is purchased with Ethereum’s native cryptocurrency, Ether, and helps to prevent spam on the network and allocate resources proportionally [28].

Smart contracts in Ethereum are typically written in a high-level programming language, such as Solidity, then compiled to EVM bytecode to be deployed to the blockchain. The EVM executes this bytecode on each node when a function from a contract is called. Due to its design, the EVM can execute untrusted code without compromising the security or performance of the network, making it a cornerstone of Ethereum’s smart contract capabilities.

E. Non-Fungible Tokens (NFTs)

Non-fungible tokens (NFTs) have gained considerable attention in the digital art and collectibles space, giving individuals the ability to prove ownership of unique pieces of content on the blockchain. In contrast to fungible tokens such as Bitcoin or Ether, NFTs are not interchangeable for other tokens of the same type but represent something unique. This uniqueness and the ability to prove ownership make NFTs particularly useful for digital art, real estate, and other use cases where uniqueness is important.

NFTs are defined in a smart contract through the ERC721 standard on the Ethereum blockchain [29]. This standard outlines a minimum interface that NFTs must implement to enable their interoperability across the Ethereum ecosystem. The ERC721 standard has given rise to many unique digital assets, from digital cats in the game CryptoKitties to multi-million dollar digital artwork.

F. Blockchain Platforms

1) *Binace Smart Chain (BSC)*: Binace Smart Chain (BSC) is a blockchain network built for running smart contract-based applications, achieving a balance between speed, security, and cost². BSC runs in parallel with Binace’s native Binace Chain (BC), hence enabling users to get the best of both worlds: the high transaction capacity of BC and the smart contract functionality of BSC.

BSC uses a consensus model called Proof of Staked Authority (PoSA), where participants stake BNB (the Binace native token) to become validators. If they propose a valid block, they’ll receive transaction fees from the transactions included in it.

BSC supports EVM, meaning that it can run Ethereum-based applications and uses tools like Metamask, Truffle, and Remix, among others. This compatibility allows it to tap into a broad developer community and existing applications, enhancing its utility and potential for adoption.

2) *Fantom*: Fantom is a high-performance, scalable, customizable, and secure smart-contract platform. It is designed to overcome the limitations of previous generation blockchain platforms³. Fantom is permissionless, decentralized, and open-source.

The primary innovation behind Fantom is a new protocol known as the “Lachesis Protocol” used to maintain consensus within the network. This protocol is intended to be highly scalable and provide near-instant transaction confirmation, making it ideal for DeFi applications and real-world uses.

Fantom is EVM-compatible, hence it allows developers to deploy Ethereum smart contracts directly to Fantom. The network uses a Proof-of-Stake (PoS) consensus algorithm and boasts high speed and low fees, offering 2-second finality for transactions.

3) *Celo*: Celo is a blockchain ecosystem focused on increasing cryptocurrency adoption among smartphone users⁴. By using phone numbers as public keys, Celo hopes to introduce the world’s billions of smartphone owners, including those without access to traditional banking services, to the benefits of cryptocurrency.

Celo’s native token is the Celo Dollar (cUSD), a stablecoin pegged to the US Dollar. This focus on a stable digital currency separates Celo from other EVM-compatible blockchains.

Celo uses a consensus mechanism called Byzantine Fault Tolerance (BFT), derived from PBFT, to maintain network security and reach consensus efficiently. It also implements an on-chain governance system that allows token holders to vote on network changes.

4) *Polygon (Matic)*: Polygon (previously Matic Network) is a Layer 2 scaling solution for Ethereum⁵. It is designed to provide faster and cheaper transactions on Ethereum using Layer 2 sidechains, which are blockchains that run alongside the Ethereum main chain. Users can deposit Ethereum tokens to a Polygon smart contract, interact with them within Polygon, and then later withdraw them back to the Ethereum main chain if necessary.

Polygon uses a modified version of the Plasma framework, an off-chain scaling solution originally proposed by Vitalik Buterin. The network also uses a Proof-of-Stake (PoS) consensus mechanism, and block producers are selected from the staking nodes.

Polygon is interoperable with a number of other blockchain networks. It supports a flexible framework for building various kinds of applications, including DeFi (Decentralized Finance) and dApps (Decentralized Applications).

The aforementioned platforms represent a selection of EVM-compatible blockchains with various features and capabilities. When deciding on a suitable platform for a particular use case, considerations such as transaction speed, cost, security, consensus mechanism, and the platform’s overall community and ecosystem need to be factored into the decision-making process.

³<https://whitepaper.io/document/438/fantom-whitepaper>

⁴<https://celo.org/papers/whitepaper>

⁵<https://polygon.technology/lightpaper-polygon.pdf>

²<https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md>

III. RELATED WORK

This section offers a comprehensive review of the previous investigations focused on the deployment of blockchain technology and smart contracts in waste management processes. To our understanding, there remains a dearth of research examining waste segregation issues within the context of a developing country. As such, this review concentrates on two primary research domains - the application of blockchain technology in managing medical waste and household waste.

A. Implementing Waste Management Models to Realize a Circular Economy (CE)

The circular economy (CE) is an aspirational model for the future that aims for sustainability through closed-loop waste management and optimal resource utilization. It is gaining attention from numerous technology companies, one of them being Amazon. The company has embarked on an initiative known as Amazon CE [30], which creates a continuous loop of product use based on partnerships and service offerings. The program empowers customers with options to reuse, repair, and recycle their products, thus aligning with the principles of the CE model.

Despite the diversity of waste types, numerous innovative strategies have emerged to handle each one. For electronic waste, Gupta et al. [31] have conceived an Ethereum-based waste management model. This model focuses on three key user groups, namely producers, consumers, and retailers, each playing a specific role in the waste management cycle. The retailers serve a dual purpose by distributing new products to consumers and collecting used ones for return to manufacturers. The correct execution of these activities rewards the participants with Ethereum's cryptocurrency, ETH.

In the context of solid waste, such as discarded computers and smartphones, Laura et al. [32] have introduced a management system founded on a combination of Ethereum and QR codes. This approach empowers stakeholders with the ability to track and ascertain the current location of waste and predict the time needed for its processing. Similarly, Schmelz et al. [33] proposed a secure and tamper-proof system for tracking cross-border waste movements using Ethereum. However, a significant drawback of this system is its inability to support penalties for waste management violations.

B. Medical Waste Management Models

The application of the CE model in a medical environment presents unique challenges. Medical equipment and supplies, which constitute a significant proportion of medical waste, are often single-use and unrecyclable after six months from their first utilization [1]. The Covid-19 pandemic has exacerbated this problem by creating an enormous quantity of medical waste, including personal protective equipment, leading to potential infection risks [34], [35].

To address these pressing issues, Trieu et al. [10] have proposed a model called MedicalWaste-Chain based on the Hyperledger Fabric. This model focuses on the treatment and disposal of medical waste emanating from health centers, as well as the recycling of tools and medical supplies. Moreover, Ahmad et al. [36] have directed their efforts toward developing

a traceability model for personal protective equipment, particularly for healthcare workers, to maintain accountability during a pandemic. To facilitate the validation of waste treatment processes and interactions between stakeholders, Dasaklis et al. [37] proposed a blockchain-based system deployable on smartphones.

C. Analysis of Blockchain Technology-based Approaches Applied to Vietnam

The approaches reviewed above, while innovative, have limitations. They tend to overlook the process of waste reproduction or refurbishment and lack comprehensive solutions for managing the behavior of end-users, particularly in terms of rewarding compliant behavior or penalizing violations. Moreover, they concentrate predominantly on managing the waste treatment chain from origin points, such as medical centers, to waste processing plants, with little consideration for household waste management.

The application and implementation of these models in a specific region like Vietnam require a holistic understanding of various socio-economic and environmental factors. As such, this study aims to instill responsible waste segregation habits among not only medical centers but also households. This research can provide a critical foundation for responding to respiratory diseases in the future, encouraging every household to adopt responsible waste disposal practices. The proposed model in this paper not only manages the waste sorting process but also incorporates a unique solution for rewarding compliant behavior and penalizing violations using Non-Fungible Token (NFT) technology. A detailed explanation of this proposed model and its implementation steps will be presented in the subsequent sections.

IV. METHODOLOGY

A. Conventional Model for Medical Waste Treatment and Classification

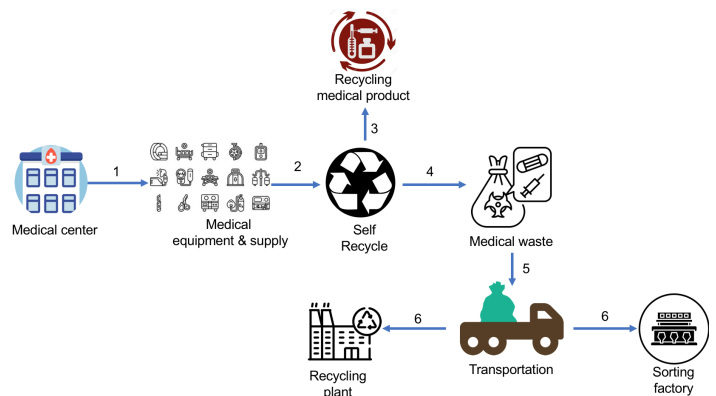


Fig. 1. Conventional model for medical waste treatment and classification.

The existing model for medical waste treatment and classification, as depicted in Fig. 1, is based on guidelines issued by the Ministry of Health in Vietnam during the Covid-19 pandemic [38]. As seen in Fig. 2, five distinct sources of medical waste are classified, which then undergo five sequential treatment steps. Medical waste is primarily generated at

treatment centers (hospitals, military barracks), testing and vaccination sites, and individual locations under quarantine (like households, apartments).

The initial three steps in medical waste classification - separation, segregation, and collection - are conducted at healthcare centers. Following this, all hazardous waste is sent to disposal facilities where it undergoes the final two steps: transportation and destruction.

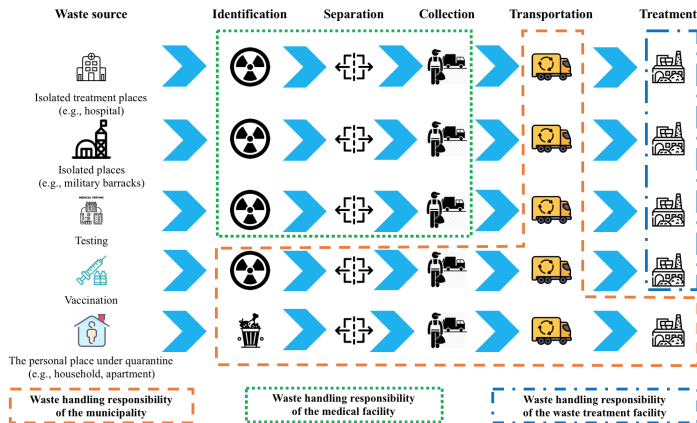


Fig. 2. Sample of medical waste treatment during the Covid-19 pandemic in Vietnam.

Under real-world circumstances, such as the care and treatment of diseases in medical centers, waste can be bifurcated into two categories: reusable and disposable. Each category warrants distinct treatment procedures.

The traditional model for waste classification and treatment, detailed in Fig. 1, is a multi-step process. Step 1 includes the collection of waste from various departments within a medical facility, primarily consisting of medical equipment and supplies. Step 2 involves the segregation of waste and identification of reusable items within the medical facility. Reusable items (Step 3) are reintegrated into the system, while disposable waste is readied for disposal (Step 4) and sent to the waste treatment area (Step 5). At this stage, waste is classified based on the requirements of treatment procedures, such as recycling or sorting (Step 6).

However, the traditional model exhibits several drawbacks, including a lack of incentive for individuals to segregate waste accurately, a deficiency of mechanisms to penalize violations, and inefficiencies in the tracking and auditing of waste treatment procedures. To address these issues, we propose a model that combines blockchain, smart contracts, and Non-Fungible Tokens (NFTs) to certify waste classification at medical centers and effectively identify compliance or violations with medical waste segregation requirements.

B. Innovative Model for Medical Waste Treatment and Classification Leveraging Blockchain Technology, Smart Contracts, and NFT

Fig. 3 illustrates the proposed model that integrates blockchain technology, smart contracts, and NFTs in a nine-step process for medical waste classification and treatment. Initially, medical professionals (doctors and nurses) familiarize

themselves with the regulations and requirements for waste segregation (Step 1). The degree of compliance with these rules becomes a crucial metric for assessing performance and determining rewards or penalties. Subsequently, medical professionals perform the initial waste segregation (self-recycling in Step 2). Hazardous waste is segregated and placed outside the patient care and treatment areas in hospitals or medical centers (Step 3). The cleaning staff, trained in assessing the waste sorting behaviors of medical personnel, conduct an initial inspection (Step 4). The inspection involves two stages, where Step 5 includes a non-invasive observation of medical staff's waste sorting activities during treatment, while Step 6 involves assessment of reusable waste in the medical environment.

Upon confirmation of compliance or violation of waste segregation requirements, the cleaning staff updates the results in the predefined functions on the smart contract (Step 7). Following this, Non-Fungible Tokens (NFTs) are generated, corresponding to the waste segregation behavior of the individuals or organizations involved (Step 8). These NFTs encapsulate relevant evidence and information concerning the individual's or organization's compliance or violation. Finally, all evaluation and validation steps, along with their results, are recorded and stored on distributed ledgers (Step 9). This blockchain-based ledger provides a transparent, secure, and immutable record of all activities, fostering accountability, and efficient auditing of medical waste management.

V. SYSTEM EXECUTION

The practical execution of our novel model is focused on two fundamental targets: i) management of medical waste data including initialization, interrogation, and modification on a blockchain platform, and ii) production of Non-Fungible Tokens (NFTs) for each user's (entities or institutions) reward and infraction behavior stemming from their participation in waste classification/disposal.

A. Data Input and NFT Initialization

The diagram in Fig. 4 details the process to initiate medical waste data. This waste includes various medical apparatus (for instance, expired or damaged) or medical consumables (such as masks, PPE, injections). These waste categories are further segregated into different classes (for example, discard, reuse) based on their toxicity grading.

Every waste bag, tagged with a unique identifier, houses a particular type of waste and carries a detailed waste description. It also encompasses metadata like the sorter's details, departmental information, time stamp, and waste generation location. The storage mechanism has been designed to handle simultaneous storage on a distributed ledger - enabling multiple users for concurrent storage to optimize system latency.

The medical waste data is structured as follows:

```
medicalWasteObject = {
  "wasteID": wasteID,
  "staffKey": staffKey,
  "category": waste category,
  "deptID": deptID,
  "amount": amount,
```

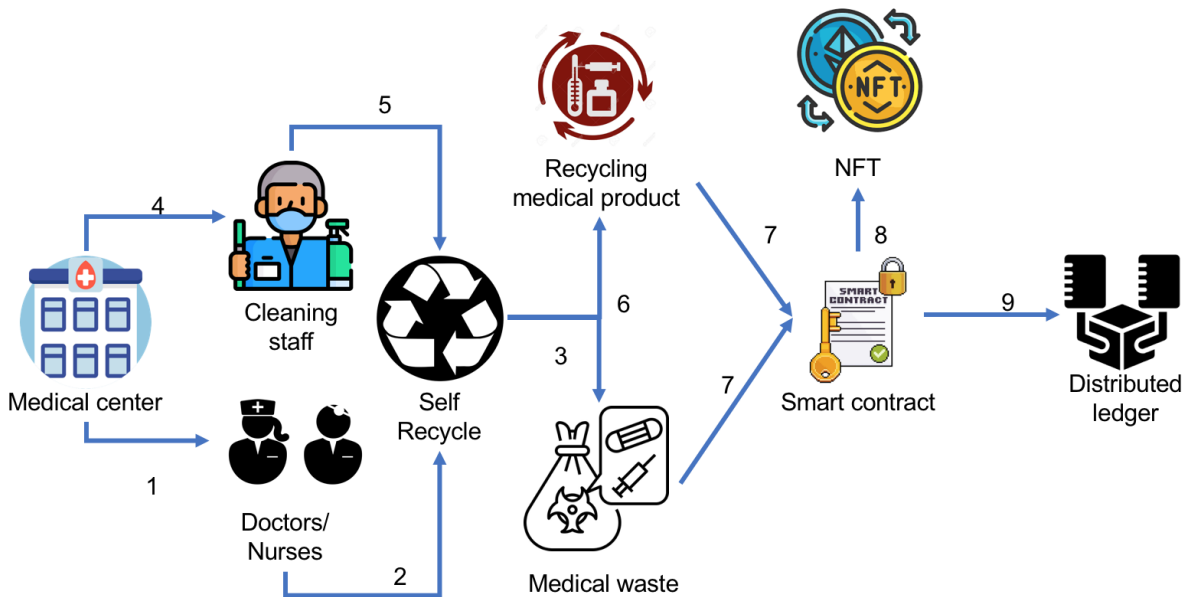


Fig. 3. Innovative model for medical waste treatment and classification leveraging blockchain technology, smart contract, and NFT.

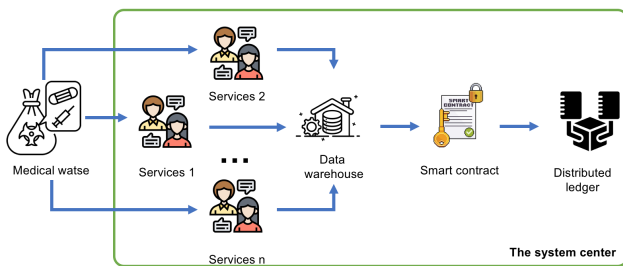


Fig. 4. Data input and NFT initialization.

```

"unitType": unitType,
"bagID": bagID,
"timestamp": timestamp,
"locale": locale,
"currentStatus": null,
"recycleStatus": Null
};

```

Alongside the essential information (such as origin, weight, waste category, etc.), we also retain information pertaining to the status of the waste bags at the medical center (“currentStatus” and “recycleStatus” - default to Null). Specifically, “currentStatus” changes to 1 if the corresponding waste bag has been dispatched out of the medical center for waste treatment; value 0 indicates a pending status. Meanwhile, “recycleStatus” becomes 1 when the waste (medical equipment) is reused (value 0 indicates pending). Non-toxic wastes pose no harm to the environment or human health.

Once the waste sorting is completed, the cleaning staff verifies the process, and upon validation, the data is synchronized onto the chain (initially stored in the data warehouse). The validation constraints embedded in the Smart Contract are activated via the API for chain synchronization. This process is crucial since it directly impacts waste treatment procedures

and forms the basis for reward or penalty for individuals and organizations.

For initiating NFTs (reward, sanction), the NFT structure is defined as:

```

NFT WASTE_HANDLING = {
"wasteID": wasteID,
"staffKey": staffKey,
"deptID": deptID,
"bagID": bagID,
"typeMatch": true/false,
"quantityMatch": true/false,
"timestamp": timestamp,
"verifierKey": staffKey // Cleaning staff
};

```

If the sorted trash bags meet the expected standards, the sorter is rewarded. If they deviate, they are penalized. The verifier is penalized in cases of incorrect information verification.

B. Data Interrogation

The data interrogation process, demonstrated in Fig. 5, is designed to support multiple simultaneous system participants. Both cleaning staff and healthcare professionals can utilize this feature, albeit for different purposes. The cleaning staff accesses data to verify the classification process or to manage the transportation of hazardous medical waste. Healthcare professionals, on the other hand, may require data to identify reusable medical tools.

These requests are submitted via API calls from the user to the system’s smart contracts, which fetch the required data from the distributed ledger. Each data retrieval request is logged as part of the query history for each individual or organization. If no match is found (e.g., incorrect ID),

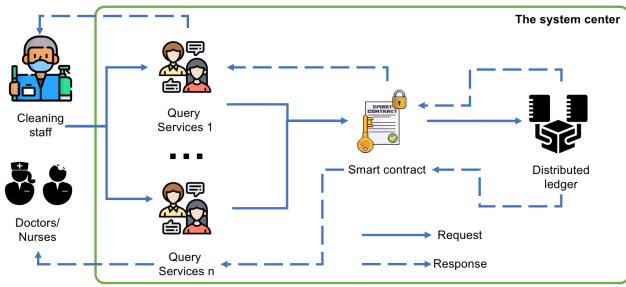


Fig. 5. Data interrogation.

the system sends an error message to the user. For NFT interrogation, APIs are provided as support services.

C. Data Modification

The data modification function, shown in Fig. 6, is activated only after data existence on the chain is confirmed. If no data is found, the system sends a corresponding error message to the user. Like data interrogation and input processes, data modification is facilitated through APIs, which process user requests and pass them onto smart contracts for execution.

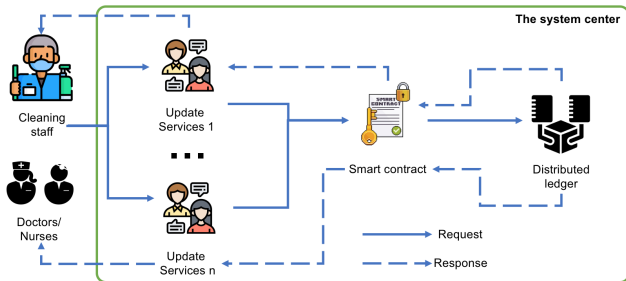


Fig. 6. Data modification.

The main objective of this function is to update the location and time stamp of waste bags during transportation and handling of medical waste. This enables administrators to trace the status of medical waste treatment/transportation from medical centers to waste treatment companies.

For NFTs, the modification process primarily involves transferring the NFT from the initial owner's address to a new one. In the event of any information update on an existing NFT, a new NFT is created.

VI. PERFORMANCE ASSESSMENT

A. Environmental Setting

The process of evaluation is a crucial aspect of our proposed model's successful deployment. It provides valuable insights about the efficiency, cost-effectiveness, and scalability of the system, particularly on Ethereum Virtual Machine (EVM)-enabled platforms. It's vital to note that the model in focus rewards or imposes penalties based on the compliance or violation of medical waste classification norms, respectively.

Earlier research publications have detailed the evaluation of system responsiveness, covering aspects like successful and

failed request responses, system latency (minimum, maximum, average), among other metrics. Hence, the focus in this current analysis pivots towards identifying the most favorable platform for deploying our proposed model.

In our attempt to identify an optimal environment for implementation, we tested our system on four renowned blockchain platforms, each boasting support for the Ethereum Virtual Machine (EVM). The chosen platforms for this comparative analysis include Binance Smart Chain (BNB Smart Chain), Polygon, Fantom, and Celo.

Fig. 7 provides a snapshot of transaction details on the Binance Smart Chain, an example of one of the four platforms evaluated. The same procedure has been repeated for all four platforms, where we successfully deployed the recommendation system and obtained transaction information. It is important to note that the transaction fees were collected in the respective native tokens of each platform. This uniform approach ensures a fair and objective evaluation process across all tested platforms.

Fig. 8 illustrates the process of NFT creation. The figure highlights how our recommendation system creates a Non-Fungible Token (NFT) as a reward or penalty mechanism for compliance or violation of waste classification norms. NFTs are created upon validation of waste sorting data by cleaning staff and are synchronized onto the chain. The data stored in the NFT includes information about the waste, staff, and department, as well as the status of type and quantity matching.

Fig. 9 showcases the process of transferring an NFT. This step is fundamental to the modification process, where the ownership of an NFT is shifted from its original holder to a new one. An essential aspect of this process is the update of the NFT ownership address.

Our performance assessment extends to smart contracts designed based on the Solidity language. These contracts were deployed in the testnet environments of all four platforms to derive a comparative analysis about the cost-effectiveness of each. Specific focus areas of our evaluation revolved around transaction fees, gas limit, gas used by the transaction, and gas price. These metrics collectively assist in identifying the most cost-effective platform for deploying our model.

By meticulously assessing these parameters and documenting the outcome, we aim to shed light on the best platform for implementing our model. Such a comparative analysis can serve as a benchmark for future implementations and modifications of the model.

B. Our Deployment in the Four Blockchain Platforms

Our evaluation encompasses four primary EVM-supported platforms, namely Binance Smart Chain, Polygon, Fantom, and Celo. For each of these platforms, we've effectively executed the deployment of our recommendation model and documented the corresponding transaction details. The relevant links that provide access to our implementation on each platform are shared below:

1) *Binance Smart Chain (BNB Smart Chain)*: This platform is an independent blockchain that runs in parallel to Binance Chain, maintaining the performance of the original

Txn Hash	Method	Block	Age	From	To	Value	[Txn Fee]
0xd74fcefb7a30f394ce9...	Transfer	24862171	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	0x94d93a5606bd3ac9ae...	0 BNB	0.00057003
0x762252a63bb7127eea...	Mint	24862162	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	0x94d93a5606bd3ac9ae...	0 BNB	0.00109162
0xf897cc7341539f38b66...	Contract Creation	24862154	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	Contract Creation	0 BNB	0.02731376

Fig. 7. Transaction details on the Binance Smart Chain.

Fig. 8. Process of NFT creation.

chain while also possessing the capability to support complex applications like decentralized apps (dApps). Our implementation of the recommendation model on the Binance Smart Chain can be viewed using the link: BNB⁶.

2) *Polygon*: A multi-chain Ethereum scaling solution that aims to provide secure, scalable, and instant transactions powered by PoS side chains. We deployed our model on the Polygon platform, and the smart contract details can be accessed at the following link: MATIC⁷.

3) *Fantom*: Known for its high-speed, low-cost transactions and secure execution of smart contracts, Fantom provides a conducive environment for deploying our recommendation model. The Fantom implementation details can be found at the subsequent link: FTM⁸.

4) *Celo*: This platform is a mobile-first platform that makes financial dApps and crypto payments accessible to anyone with a mobile phone. Our model’s deployment on the Celo platform can be examined at the subsequent link: CELO⁹.

Each of these links directs the user to the respective testnet environments where the detailed implementation of our recommendation model on each platform is available. The information presented includes an overview of the transactions associated with our smart contracts, including transaction hash, status, block, timestamp, from, to, value, and transaction fee, among others. It provides a comprehensive snapshot of the

⁶<https://testnet.bscscan.com/address/0x94d93a5606bd3ac9ae8b80e334dfec74d0075ece>

⁷<https://mumbai.polygonscan.com/address/0x48493a3bb4e7cb42269062957bd541d52afc0d7a>

⁸<https://testnet.ftmscan.com/address/0x48493a3bb4e7cb42269062957bd541d52afc0d7a>

⁹<https://explorer.celo.org/alfajores/address/0x48493A3bB4E7cB42269062957Bd541D52aFc0d7A/transactions>

process of deploying our model and the associated costs for each of the four platforms.

C. Transaction Fee

Table I, titled “Transaction fee”, presents a comparative overview of the transaction fees associated with various operations conducted on four distinct blockchain platforms: Binance Smart Chain (BNB), Fantom, Polygon (MATIC), and Celo.

The operations encapsulate:

- **Contract Creation**: This signifies the process of deploying a novel smart contract onto the blockchain network. A smart contract represents a self-executing contract with the agreement terms being inscribed directly into the code, subsequently stored and replicated on the blockchain.
- **Create NFT**: This operation encompasses the generation of a Non-Fungible Token (NFT) on the blockchain. NFTs constitute a genre of digital asset created to showcase ownership or authentication proof of unique items or content.
- **Transfer NFT**: This operation details the procedure of transferring the ownership of an NFT from one entity to another within the blockchain.

For every operation, the transaction fees are delineated for each blockchain platform in their specific cryptocurrency (BNB, FTM, MATIC, CELO), alongside their equivalent value in USD (\$) within brackets.

- For Binance Smart Chain (BNB), the costs for Contract Creation, Create NFT, and Transfer NFT operations are 0.02731376 BNB (\$8.41), 0.00109162 BNB (\$0.34), and 0.00057003 BNB (\$0.18) respectively.
- For Fantom, the corresponding costs stand at 0.009577666 FTM (\$0.001840), 0.000405167 FTM (\$0.000078), and 0.0002380105 FTM (\$0.000046).
- For Polygon (MATIC), the costs are calculated as 0.006841190030101236 MATIC (\$0.01), 0.000289405001041858 MATIC (\$0.00), and 0.000170007500612027 MATIC (\$0.00).
- Lastly, for Celo, the costs are evaluated as 0.0070979376 CELO (\$0.004), 0.0002840812 CELO (\$0.000), and 0.0001554878 CELO (\$0.000).

Txn Hash	Age	From	To	Token ID	Token
0xd74fcefb7a30f394ce9...	1 day 22 hrs ago	0x94d93a5606bd3ac9ae...	OUT 0xcaa9c5b45206e083f4f...	1	ERC-721: NFT....ENT
0x762252a63bb7127eea...	1 day 22 hrs ago	0x000000000000000000...	IN 0x94d93a5606bd3ac9ae...	1	ERC-721: NFT....ENT

Fig. 9. NFT transfer process.

TABLE I. TRANSACTION FEE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.02731376 BNB (\$8.41)	0.00109162 BNB (\$0.34)	0.00057003 BNB (\$0.18)
Fantom	0.009577666 FTM (\$0.001840)	0.000405167 FTM (\$0.000078)	0.0002380105 FTM (\$0.000046)
Polygon	0.006841190030101236 MATIC(\$0.01)	0.000289405001041858 MATIC(\$0.00)	0.000170007500612027 MATIC(\$0.00)
Celo	0.0070979376 CELO (\$0.004)	0.0002840812 CELO (\$0.000)	0.0001554878 CELO (\$0.000)

TABLE II. GAS LIMIT

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,376	109,162	72,003
Fantom	2,736,476	115,762	72,803
Polygon	2,736,476	115,762	72,803
Celo	3,548,968	142,040	85,673

TABLE III. GAS USED BY TRANSACTION

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,376 (100%)	109,162 (100%)	57,003 (79.17%)
Fantom	2,736,476 (100%)	115,762 (100%)	68,003 (93.41%)
Polygon	2,736,476 (100%)	115,762 (100%)	68,003 (93.41%)
Celo	2,729,976 (76.92%)	109,262 (76.92%)	59,803 (69.8%)

This table aids in visualizing and contrasting the cost-effectiveness of deploying and managing NFTs across these platforms, thereby facilitating the decision-making process for selecting the most fitting blockchain for specific applications.

D. Gas limit

In the realm of blockchain technology, the term “Gas Limit” carries a specific significance. The gas limit can be understood as the maximum amount of computational power an individual is willing to expend for conducting a particular operation or executing a transaction on the blockchain. In essence, it acts as a cap to prevent overspending or infinite looping of operations. Since every operation, from simple to complex, on the blockchain requires a certain amount of computational resources, the gas limit ensures that these operations do not overrun their resource allocation.

Operations such as contract creation, NFT creation, or NFT transfer all necessitate different amounts of computational resources, hence different gas limits. The gas limit is explicitly set for each transaction, and if an operation exceeds this set limit, it will be terminated, ensuring the integrity of the network and the safety of its users.

In the context of the Table II, titled “Gas limit”, the values detailed represent the gas limits for different operations, namely contract creation, NFT creation, and NFT transfer across four blockchain platforms: Binance Smart Chain (BNB), Fantom, Polygon (MATIC), and Celo. The gas limits are presented in units of gas.

For the Binance Smart Chain, the gas limits for contract creation, NFT creation, and NFT transfer are set at 2,731,376, 109,162, and 72,003 gas units, respectively.

Fantom and Polygon share identical gas limit values, with 2,736,476 units for contract creation, 115,762 units for NFT creation, and 72,803 units for NFT transfer.

On the other hand, Celo requires the highest amount of gas for each operation. The gas limits for contract creation, NFT creation, and NFT transfer on Celo are 3,548,968, 142,040, and 85,673 units, respectively.

This comparative analysis of gas limits across different platforms aids in assessing the computational efficiency of conducting operations on these platforms. It provides crucial insights into the operational costs involved in the deployment and management of smart contracts and NFTs, which can guide the selection of the most appropriate and cost-effective platform for specific use cases.

E. Gas Used by Transaction

In the context of blockchain transactions, “Gas Used by Transaction” refers to the actual amount of computational work done by a particular transaction on the blockchain network. Each operation or instruction in a transaction requires a certain amount of gas to execute, and the total gas used by the transaction is the sum of the gas used by each of these individual operations.

It’s crucial to understand that not all set gas (defined by the gas limit) is always consumed by a transaction. The actual gas consumed depends on the computational complexity of the transaction. If a transaction finishes before reaching its gas limit, the unused gas is refunded to the sender. Conversely, if a transaction runs out of gas before it completes, it is halted, and all changes are reversed, but no gas is returned. Therefore, the gas used by transaction metric can provide an insight into the computational efficiency and cost-effectiveness of a transaction.

In Table III, titled “Gas Used by Transaction”, we present the actual gas used by three different types of transactions — contract creation, NFT creation, and NFT transfer — on four different blockchain platforms: Binance Smart Chain (BNB),

Fantom, Polygon (MATIC), and Celo. The values are reported in units of gas and also as a percentage of the respective gas limit for each transaction type.

For contract creation and NFT creation transactions, BNB, Fantom, and Polygon all use 100% of the gas limit, indicating that these transactions use all allocated resources. For NFT transfer, BNB uses 79.17% of the gas limit, while Fantom and Polygon use slightly more, at 93.41%.

Interestingly, Celo exhibits a different pattern. For contract creation and NFT creation transactions, Celo uses only 76.92% of the gas limit. This indicates a higher computational efficiency for these types of transactions on Celo compared to the other platforms. However, the NFT transfer on Celo uses only 69.8% of the gas limit, which is lower than the corresponding values on the other platforms.

This comparative study provides a clear picture of the computational efficiency of executing different transactions across various platforms. These insights can significantly aid in selecting the most efficient and cost-effective platform for deploying and managing smart contracts and NFTs.

F. Gas Price

In blockchain ecosystems, “Gas Price” represents the cost of each unit of gas that a user is willing to pay for a transaction. The unit for gas price is typically “gwei” (giga-wei), where 1 ETH (Ether) equals 1,000,000,000 gwei. The gas price is set by the sender of the transaction and plays a significant role in transaction prioritization. Miners, who validate and add transactions to the blockchain, have a preference for transactions with higher gas prices, as it leads to greater rewards for them. Consequently, if a user sets a higher gas price, their transaction is likely to be processed more quickly. However, setting an exceedingly high gas price can lead to unnecessary costs, while setting it too low might result in the transaction not getting processed if miners deem it unworthy of their computational effort. Therefore, users need to find a balance to ensure that their transactions are processed in a reasonable timeframe without incurring excessive costs.

In Table IV, titled “Gas Price”, we provide a detailed comparison of the gas prices for three different types of transactions — contract creation, NFT creation, and NFT transfer — across four different blockchain platforms: Binance Smart Chain (BNB), Fantom, Polygon (MATIC), and Celo.

For BNB Smart Chain, the gas price for all three transaction types is set at 0.00000001 BNB, equivalent to 10 gwei. This is a common gas price on the BNB Smart Chain and is likely to ensure a swift transaction execution. On the Fantom network, the gas price is lower at 0.0000000035 FTM, or 3.5 gwei for all three transactions. This reduced price could result in slower transaction processing times, but it also means lower transaction costs. For Polygon, the gas price for all transaction types is set even lower at 0.000000002500000011 MATIC, approximately equivalent to 2.5 gwei. Again, this could potentially lead to slower transaction times but lower costs. Finally, for Celo, the gas price for all transactions is set at 0.0000000026 CELO. Notably, this platform also specifies a “Max Fee per Gas” set at 2.7 Gwei. This is the maximum price that the sender is willing to pay per unit of gas, which gives the user more control over the transaction costs.

This comprehensive comparison across various platforms can help users to make informed decisions when choosing the optimal platform for their specific needs, taking into account both transaction costs and expected processing times.

VII. DISCUSSION

A. Threats to Validity

The evaluation carried out in this research attempts to measure and compare the performance of various blockchain platforms, focusing on transaction costs, response rates, and other metrics. However, several potential threats to validity need to be acknowledged for a comprehensive understanding of the findings.

The first and foremost issue pertains to the inherently volatile nature of cryptocurrency markets. The conversion rates and transaction costs cited in this study represent a snapshot of market conditions at a specific point in time, and do not account for the periodic and often substantial fluctuations that can drastically affect these figures. Therefore, the calculated cost-effectiveness of each platform, as presented in this research, may vary significantly depending on the market state at the time of consultation.

Secondly, the study assumes a controlled environment for all platforms without any network congestion, excessive transaction volumes, or other real-time factors that could potentially influence the performance and responsiveness of a platform. These uncontrollable real-world variables can yield different results under varying circumstances, thus impacting the generalizability of the study’s conclusions.

Finally, the evaluation was conducted on the testnet environments of the four platforms, which might not fully represent the conditions of the main networks. The responsiveness and performance on the mainnet could differ, affecting the validity of the comparisons made in this research.

B. Notable Observations

In the process of conducting this comparative study, several notable observations were made. The Binance Smart Chain (BNB) demonstrated the highest transaction costs among the four platforms evaluated. However, it also consistently provided high transaction throughput, which may be crucial for applications requiring rapid, high-volume transactions.

On the other hand, platforms such as Fantom, Polygon, and Celo displayed significantly lower transaction costs, which can be an attractive attribute for applications sensitive to cost constraints. Nonetheless, the lower costs may correspond to a slower transaction speed due to decreased miner incentives. These variations underline the trade-offs that need to be considered when choosing a blockchain platform for application deployment.

C. Limitations

This research was conducted with certain limitations which should be taken into account while interpreting the findings. Primarily, the implementations were executed in controlled test environments, which may not replicate the real-world

TABLE IV. GAS PRICE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)
Fantom	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)	0.0000000035 FTM (3.5 Gwei)
Polygon	0.000000002500000011 MATIC (2.500000011 Gwei)	0.000000002500000009 MATIC (2.500000009 Gwei)	0.000000002500000009 MATIC (2.500000009 Gwei)
Celo	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.0000000026 CELO (Max Fee per Gas: 2.7 Gwei)

conditions of live networks. Variables such as network congestion, transaction volume, and changing miner incentives can significantly influence the transaction fees, gas usage, and response times experienced on the live networks.

Secondly, the study's focus is largely technical and quantitative, revolving around performance metrics and cost factors. It does not consider qualitative aspects such as ease of use, community support, or developer tools provided by the platforms, which can also influence the selection of a blockchain platform.

Furthermore, the study did not account for potential changes in the platforms themselves. Modifications or updates to the platforms' protocols, changes in the consensus mechanisms, or introduction of new features could significantly alter the performance or cost structure, rendering the current findings less relevant.

D. Future Work

Building upon this research, future studies could encompass a wider array of blockchain platforms for a more comprehensive comparison. Moreover, evaluations could be conducted under varying network conditions to capture a more accurate picture of how factors such as network congestion or increased transaction volumes affect transaction costs and performance.

Additionally, a more holistic approach could be adopted to consider qualitative aspects in addition to the technical and quantitative parameters evaluated in this study. These could include ease of use, developer support, platform maturity, and other factors that could influence the choice of a platform.

Future research could also closely monitor updates and modifications to these blockchain platforms, in order to assess how these changes affect the performance and cost effectiveness. Lastly, a deeper investigation into the security aspects of these platforms could be carried out. This is particularly important as security is a key consideration in blockchain applications, and this aspect was not the main focus of the current study.

In our future research plans, we also intend to venture further into the development and integration of sophisticated algorithms, with a particular focus on encryption and decryption methodologies. These strategies provide an additional layer of security to our model, ensuring the privacy and confidentiality of data transactions on the blockchain. More specifically, we aim to examine the transactional costs associated with implementing such complex methodologies. We hope to elucidate the correlation between the complexity of data structures and transaction costs, thus providing a more comprehensive picture of cost-effectiveness in blockchain deployment.

Simultaneously, we are exploring the idea of implementing our proposed model in a live, real-world environment. While our initial studies have taken place in controlled, simulated scenarios, a deployment on a mainnet environment such as Fantom (FTM) will expose our system to real-world dynamics. This could offer valuable insights into the practicalities of deploying blockchain systems, and the unique challenges that might arise therein.

Our current analysis, while comprehensive, does not yet fully consider the nuances of user privacy policies. Access control, a critical aspect of any system dealing with user data, has been examined in previous studies [39], [40]. Similarly, dynamic policies, which allow for flexibility and adaptability in system rules, have also been the focus of earlier research [41], [42]. In our forthcoming research activities, we plan to delve into these areas. Our aim is to establish a robust privacy framework that strikes a balance between data security and operational efficiency.

In terms of infrastructure, we are looking at the possibility of incorporating certain proven approaches into our model. Technologies such as gRPC [43], [44], a high-performance remote procedure call (RPC) framework, offer benefits in terms of speed and interoperability. Microservices architecture, too, presents a scalable and efficient way to structure applications [45], [46]. Similarly, dynamic message transmission strategies [47] and brokerless systems [48] have their respective advantages in enhancing user interaction. Incorporating these technologies via an API-call-based approach could create a more intuitive, accessible, and efficient system for users interacting with the blockchain.

VIII. CONCLUSION

In summation, this study has bestowed invaluable understanding pertaining to the selection of appropriate EVM-compatible blockchain platforms for the deployment of our proposed recommendation model. Through an exhaustive investigation and assessment of platforms including Binance Smart Chain, Polygon, Fantom, and Celo, we have unearthed detailed distinctions in costs, gas limits, gas consumption, and gas prices, each playing a critical role in the creation and transfer of Non-Fungible Tokens (NFTs) and smart contract deployments.

The comprehensive evaluation of transactional expenses, gas thresholds, gas consumed, and gas pricing has not only delivered a lucid understanding of operational expenditure associated with each platform, but also unveiled the intricacies of transactional efficiency and efficacy. Binance Smart Chain emerged as a cost-effective solution, while Fantom showed promising transactional speed and effectiveness.

Our contribution extends beyond proposing a novel recommendation model, as we have openly shared the implementation details on these blockchain platforms. This initiative is expected to stimulate further research and offer the developer community an in-depth practical insight into working with these platforms. Furthermore, we have offered an elaborate account of our evaluation procedure, ensuring its reproducibility and transparency.

Our future work is ripe with exciting opportunities, from delving into complex methodologies such as encryption and decryption, addressing privacy policy issues, and investigating infrastructure-based strategies. As we incessantly refine and augment our model, these areas will form the epicenter of our research focus.

Even though the road ahead is laden with complexities, the study reiterates the enormous potential and versatility of blockchain technology across varied applications. As we steer forward, our objective is to leverage these unique strengths to craft an efficient, robust, and secure blockchain-powered recommendation system. The exploratory journey continues, with each stride taking us closer to our ultimate goal: an equitable, secure, and accessible future propelled by blockchain technology.

ACKNOWLEDGMENT

This research project received invaluable contributions from Engineer Le Thanh Tuan and Dr. Ha Xuan Son, who offered guidance and support during the brainstorming, implementation, and evaluation stages. We are also indebted to FPT University Cantho Campus, Vietnam, for their supportive role in this study.

REFERENCES

- [1] A. Leonard, *The story of stuff: How our obsession with stuff is trashing the planet, our communities, and our health—and a vision for change*. Simon and Schuster, 2010.
- [2] S. T. Wafula, J. Musiime, and F. Oporia, “Health care waste management among health workers and associated factors in primary health care facilities in kampala city, uganda: a cross-sectional study,” *BMC public health*, vol. 19, no. 1, pp. 1–10, 2019.
- [3] J. M. Turner and L. M. Nugent, “Charging up battery recycling policies: extended producer responsibility for single-use batteries in the european union, canada, and the united states,” *Journal of Industrial Ecology*, vol. 20, no. 5, pp. 1148–1158, 2016.
- [4] K. Bakhsh, S. Rose, M. F. Ali, N. Ahmad, and M. Shahbaz, “Economic growth, co2 emissions, renewable waste and fdi relation in pakistan: New evidences from 3sls,” *Journal of environmental management*, vol. 196, pp. 627–632, 2017.
- [5] N. Gaur, K. Narasimhulu, and Y. PydiSetty, “Recent advances in the bio-remediation of persistent organic pollutants and its effect on environment,” *Journal of cleaner production*, vol. 198, pp. 1602–1631, 2018.
- [6] A. K. Awasthi, X. Zeng, and J. Li, “Environmental pollution of electronic waste recycling in india: A critical review,” *Environmental pollution*, vol. 211, pp. 259–270, 2016.
- [7] F. Echegaray and F. V. Hansstein, “Assessing the intention-behavior gap in electronic waste recycling: the case of brazil,” *Journal of Cleaner Production*, vol. 142, pp. 180–190, 2017.
- [8] N. U. Benson, O. H. Fred-Ahmadu, D. E. Bassey, and A. A. Atayero, “Covid-19 pandemic and emerging plastic-based personal protective equipment waste pollution and management in africa,” *Journal of environmental chemical engineering*, vol. 9, no. 3, p. 105222, 2021.
- [9] Z. Chen, M. A. Sidell, B. Z. Huang, T. Chow, M. P. Martinez, F. Lurmann, F. D. Gilliland, and A. H. Xiang, “The independent effect of covid-19 vaccinations and air pollution exposure on risk of covid-19 hospitalizations in southern california,” *American Journal of Respiratory and Critical Care Medicine*, no. ja, 2022.
- [10] H. T. Le *et al.*, “Medical-waste chain: A medical waste collection, classification and treatment management by blockchain technology,” *Computers*, vol. 11, no. 7, p. 113, 2022.
- [11] J. Li and M. Kassem, “Applications of distributed ledger technology (dlt) and blockchain-enabled smart contracts in construction,” *Automation in construction*, vol. 132, p. 103955, 2021.
- [12] A. K. Das, M. Islam, M. Billah, A. Sarker *et al.*, “Covid-19 and municipal solid waste (msw) management: a review,” *Environmental Science and Pollution Research*, vol. 28, no. 23, pp. 28 993–29 008, 2021.
- [13] H. X. Son and E. Chen, “Towards a fine-grained access control mechanism for privacy protection and policy conflict resolution,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 2, 2019.
- [14] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” *Decentralized Business Review*, p. 21260, 2008.
- [15] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, “Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems,” in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 105–109.
- [16] —, “On components of a patient-centered healthcare system using smart contract,” in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*. New York, NY, USA: Association for Computing Machinery, 2020, p. 31–35.
- [17] X. S. Ha, H. T. Le, N. Metoui, and N. Duong-Trung, “Dem-cod: Novel access-control-based cash on delivery mechanism for decentralized marketplace,” in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2020, pp. 71–78.
- [18] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, “Multi-sessions mechanism for decentralized cash on delivery system,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019.
- [19] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, “Scrutinizing trust and transparency in cash on delivery systems,” in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.
- [20] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, “Bloodchain: A blood donation network managed by blockchain technologies,” *Network*, vol. 2, no. 1, pp. 21–35, 2022.
- [21] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, “Toward a design of blood donation management by blockchain technologies,” in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 78–90.
- [22] H. X. Son, T. H. Le, N. T. T. Quynh, H. N. D. Huy, N. Duong-Trung, and H. H. Luong, “Toward a blockchain-based technology in dealing with emergencies in patient-centered healthcare systems,” in *International Conference on Mobile, Secure, and Programmable Networking*. Springer, 2020, pp. 44–56.
- [23] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, “Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies,” in *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer, 2022, pp. 576–583.
- [24] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, “Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.

- [25] H. T. Le, N. T. T. Le, N. N. Phien, and N. Duong-Trung, "Introducing multi shippers mechanism for decentralized cash on delivery system," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.
- [26] Z. Zheng, S. Xie, H.-N. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," *Big Data Research*, vol. 2, pp. 57–93, 2020.
- [27] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum Project Yellow Paper*, 2014.
- [28] K. L. Quoc *et al.*, "Sssb: An approach to insurance for cross-border exchange by using smart contracts," in *Mobile Web and Intelligent Information Systems: 18th International Conference*. Springer, 2022, pp. 179–192.
- [29] W. Entriken *et al.* (2018) Erc721 non-fungible token standard. [Online]. Available: <https://eips.ethereum.org/EIPS/eip-721>
- [30] "How amazon is investing in a circular economy," <https://www.aboutamazon.com/news/sustainability/how-amazon-is-investing-in-a-circular-economy>, accessed: 2022-10-30.
- [31] N. Gupta and P. Bedi, "E-waste management using blockchain based smart contracts," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2018, pp. 915–921.
- [32] M. R. Laouar, Z. T. Hamad, and S. Eom, "Towards blockchain-based urban planning: Application for waste collection management," in *Proceedings of the 9th International Conference on Information Systems and Technologies*, 2019, pp. 1–6.
- [33] D. Schmelz, K. Pinter, S. Strobl, L. Zhu, P. Niemeier, and T. Grechenig, "Technical mechanics of a trans-border waste flow tracking solution based on blockchain technology," in *2019 IEEE 35th international conference on data engineering workshops (ICDEW)*. IEEE, 2019, pp. 31–36.
- [34] J. R. Sheehan, B. Lyons, and F. Holt, "The use of lean methodology to reduce personal protective equipment wastage in children undergoing congenital cardiac surgery, during the covid-19 pandemic," *Pediatric Anesthesia*, vol. 31, no. 2, pp. 213–220, 2021.
- [35] K. Manninen, S. Koskela, R. Antikainen, N. Bocken, H. Dahlbo, and A. Aminoff, "Do circular economy business models capture intended environmental value propositions?" *Journal of Cleaner Production*, vol. 171, pp. 413–422, 2018.
- [36] R. W. Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, M. Omar, and S. Ellahham, "Blockchain-based forward supply chain and waste management for covid-19 medical equipment and supplies," *Ieee Access*, vol. 9, pp. 44 905–44 927, 2021.
- [37] T. K. Dasaklis, F. Casino, and C. Patsakis, "A traceability and auditing framework for electronic equipment reverse logistics based on blockchain: the case of mobile phones," in *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, 2020, pp. 1–7.
- [38] T. D. Nguyen, K. Kawai, and T. Nakakubo, "Estimation of covid-19 waste generation and composition in vietnam for pandemic management," *Waste Management & Research*, vol. 39, no. 11, pp. 1356–1364, 2021.
- [39] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward a privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on European Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.
- [40] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.
- [41] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.
- [42] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.
- [43] L. T. T. Nguyen *et al.*, "Bmdd: a novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *PeerJ Computer Science*, vol. 8, p. e950, 2022.
- [44] L. N. T. Thanh *et al.*, "Toward a security iot platform with high rate transmission and low energy consumption," in *International Conference on Computational Science and its Applications*. Springer, 2021.
- [45] —, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management*. Springer, 2021.
- [46] L. N. T. Thanh, N. N. Phien, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2021.0120768>
- [47] L. N. T. Thanh *et al.*, "Uip2sop: A unique iot network applying single sign-on and message queue protocol," *IJACSA*, vol. 12, no. 6, 2021.
- [48] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," 2021.

A Novel Dual Confusion and Diffusion Approach for Grey Image Encryption using Multiple Chaotic Maps

S Phani Praveen¹, Dr V Sathiya Suntharam², Dr S Ravi³,
U.Harita⁴, Venkata Nagaraju Thatha⁵, D Swapna⁶

Department of Computer Science and Engineering

Prasad V Potluri Siddhartha Institute of Technology, Andhra Pradesh, India¹

Department of Computer Science and Engineering (Cyber Security), CMR Engineering College, Hyderabad, India²

Department of Electronics and Communication Engineering

Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India³

Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, A.P India⁴

Department of Information Technology

MLR Institute of Technology, Hyderabad, Telangana, India⁵

Department of Computer Science and Engineering

Gitam School of Technology, GITAM (Deemed to be University), Andhra Pradesh, India⁶

Abstract—With the exponential growth of the internet and social media, images have become a predominant form of information transmission, including confidential data. Ensuring the proper security of these images has become crucial in today's digital age. This research study proposes a unique strategy for solving this demand by presenting a dual confusion and diffusion technique for encrypting gray-scale pictures. This method is presented as an innovative means of meeting this need. To improve the effectiveness of the encryption process, the encryption method uses several chaotic maps, including the logistic map, the tent map, and the Lorenz attractor. Python is used for the implementation of the suggested approach. Furthermore, a thorough assessment of the encryption mechanism is carried out to determine its efficacy and resilience. By employing the combined strength of chaotic maps and dual confusion and diffusion techniques, the proposed method aims to provide a high level of security for confidential image transmission. The experimental results demonstrate the algorithm's effectiveness in terms of encryption speed, security, and resistance against common attacks. The encrypted images exhibit properties such as randomness, key sensitivity, and resilience against statistical analysis and differential attacks. Moreover, the proposed method maintains a reasonable computational efficiency, and it is compatible with real-time applications. This study makes a contribution to the growing area of picture encryption by presenting an original and effective encryption method that overcomes the shortcomings of previously used approaches. Future work can explore additional security features and extend the proposed approach to encrypt other forms of multimedia data.

Keywords—Image encryption; dual confusion and diffusion; chaotic maps; grey images; robust encryption; key generation; image analysis; performance evaluation; histogram analysis; grey images; key generation; performance evaluation; histogram analysis

I. INTRODUCTION

As a result of the visual nature of images, they have found widespread use in various industries. Because phone terminals may be stolen in open environments and because images in the phone terminals might be lost, and contain

enormous volumes of private information, the privacy of the information contained in photographs is at a significant risk of being compromised. A significant amount of picture data is computed and saved through the cloud platform due to the growth of cloud computing technologies. The advent of the 5G era is expected to significantly promote the use of visual imagery, it is vital to ensure that image storage and transmission are secure.

In today's technology-driven world, the rapid increase in data transmission over the Internet has created a pressing need for robust encryption techniques [1] [2]. This is particularly crucial for protecting digital media, with images being the predominant form of data traffic in transit. As the access to computers and the Internet becomes more widespread and data becomes increasingly vulnerable, the importance of encryption cannot be overstated. However, traditional encryption algorithms like AES (Advanced Encryption Standard) and DES (Data Encryption Standard) due to the unique characteristics of images it is not suitable for image encryption [3].[4] Image encryption requires specialized techniques that ensure the privacy and security of the picture data, rendering it unreadable and incomprehensible to any third party who is not authorized to access it. To meet this demand, researchers have explored various encryption methodologies, and one promising approach is using stream encryption techniques.

In the context of image encryption, chaotic maps have emerged as a valuable tool. Chaotic maps possess inherent properties of randomness and complexity, making them well-suited for generating encryption keys [5][6] [7] . Key creation using chaotic maps improves encryption system security. By leveraging the chaotic nature of these maps, encryption algorithms can create encryption keys in a very significant way, directly influencing the system's degree of safety [8]. Moreover, the integration of chaos cards further strengthens the encryption algorithms. Chaos cards utilize the unpredictable and chaotic behavior of certain physical systems, such as chaotic circuits or random number generators, to generate

highly random and secure encryption keys. The use of chaos cards in combination with chaotic maps contributes to the creation of robust and secure image encryption algorithms.

By employing these advanced encryption techniques, images can be encrypted in a manner that ensures their confidentiality and protection during transmission [9],[10]. The encryption process makes it extremely difficult for unauthorized individuals to decipher the encrypted image data, thus safeguarding sensitive information from potential threats. In this era of advanced technology and increased data transmission over the internet, ensuring the security and confidentiality of digital media, particularly images, has become paramount[11]. Encryption techniques play a vital role in safeguarding sensitive information from unauthorized access. Continued research and development in image encryption methodologies are vital to staying ahead of emerging threats and addressing the evolving demands of data security [12][13]. Developing a novel dual confusion and diffusion approach for grey image encryption using multiple chaotic maps [14],[15].

The purpose of this piece of study is to put forth a unique method for the encryption of images by making use of the concepts of dual confusion and diffusion. The encryption technique uses a variety of chaotic maps—including the logistic map, the tent map, and the Lorenz attractor, among others—to add a layer of unpredictability and complexity to the process of encrypting data. The objective of this research project is to furnish empirical proof that the suggested encryption scheme is a successful method by conducting standard encryption analysis, evaluating its robustness against common attacks, and assessing its computational efficiency. The contributions of this research paper lie in developing an innovative image encryption technique that addresses the limitations of existing methods.

- Showing the encryption algorithm's security and resilience.
- Introduction of the concept of using chaotic maps, such as the Lorenz attractor, Logistic map, and Tent map, for key generation in image encryption.
- Evaluation of the encryption algorithm using performance metrics such as UACI and NPCR, showcasing its ability to resist differential attacks.
- Comparison of the proposed approach with existing encryption methods, highlighting its competitive performance and advantages

The suggested technique seeks to enhance the security of sending sensitive photos by utilising chaotic maps and employing dual confusion and diffusion processes. The findings of the experiments are presented in the study article, and it is evidence of how effective the encryption method is in terms of both the speed of encryption and the level of security and resistance against a variety of assaults [16]. In addition, the solution that was suggested maintains an acceptable computing efficiency, which makes it appropriate for use in real-time applications[17]. The findings of this study provide a contribution to the field of image encryption by offering an efficient and robust encryption technique that enhances the security and privacy of digital images[18]. Future work in this area can

explore additional security features and extend the proposed approach to encrypt other forms of multimedia data.

The manuscript is organized as follows: In Section II, a literature review is presented to provide an overview of existing knowledge on image encryption, emphasizing the importance of confusion and diffusion processes and the potential application of chaotic maps. Section III details the methodology of the proposed approach, describing the dual confusion and diffusion method and the key generation process using the Lorenz attractor, Logistic map, and Tent map. Section IV gives the analysis methods. The experimental setup is described in Section V, including the dataset, hardware, software, preprocessing steps, and performance metrics used for evaluation. Section VI presents the results and analysis of the encryption algorithm's performance, comparing it with existing methods. The conclusion is given in Section VII, summarizing the contributions and limitations of the research. Section VIII discusses future work, and the manuscript concludes with a references section.

II. LITERATURE REVIEW

Talhaoui et al. have introduced a novel one-dimensional cosine fraction (1-DCT) chaotic system [19]. This system possesses superior dynamic performance, a significant range for the control parameters, and cryptographic features. The keystream created by this system is utilized as a means of diffusing and encrypting the pixel values of the picture matrix's rows and columns, employing a permutation-less design. This is done to acquire the cipher text image. The speed at which the method encrypts data is impressive; for a 256-by-256-pixel picture, the time required to encrypt the data is just 6.7 seconds. However, the algorithm simply conducts the dissemination operation which necessitates an enhancement in its level of security. Third-order fractional chaotic systems are proposed by Xu et al. [20] for their superior dynamical performance and expansive key spaces. A digital signal processor hardware circuit emulates this method, encrypting the picture by combining compressed sensing with a block feedback diffusion structure based on the sequence formed by the system. He also proposed that this system could be used to encrypt text. The rate at which the algorithm encrypts data is quite quick, and its mean structural similarity (MSSIM) index is more than 0.9; yet, its capacity to withstand attacks is rather poor. Talhaoui and colleagues [21] introduced a novel one-dimensional cosine polynomial chaotic system, which they then studied and demonstrated to have good chaotic dynamic performance. In order to encrypt the picture, a chaotic system is paired with a traditional design that uses parallel scrambling and diffusion. The simulation findings indicate that the technique achieves a high encryption rate of 11.1 seconds per picture with a dimension of 256 pixels by 256 pixels. Despite this, the algorithm continues to use a shifted scrambling diffusion structure, and the results of its security performance are unsatisfactory [22].

Aparna et al. [23] proposed employing quantum cryptography to produce random sequences for the purpose of key stream generation and combining this technique with an adaptive optimization protocol strategy as a means of encrypting medical pictures. Quantum cryptography was used to complete this task successfully. Although the technique exhibits the

ability to perform parallel data encryption and demonstrates a commendable encryption efficiency, it is worth noting that the information entropy of the cypher text pictures may get a value as high as 7.9974. Consequently, this contributes to a substantial level of security. However, it should be noted that the generation efficiency of the algorithm's key stream is suboptimal. Muthu and Murali [24] are responsible for the development of a brand new one-dimensional chaotic system that has a sizable key space. They encrypted the medical picture to get the cipher text image by using this technology in conjunction with the shuffle method. Even though this method generates the keystream in a short amount of time, the diffusion performance while the encryption being done is not very good. Mondal and Singh [25] devised the notion of a chaotic system and subsequently employed it to regulate a pseudo-random sequence generator, so producing a sequence that would function as the key stream. This innovation was Lightweight. Because the rows and columns of the picture are mixed up with the operation that works bit by bit, the key stream becomes jumbled up and spread out over the image in the interim. This is because the operation works bit by bit. Both the direct bit-by-bit operation that the technique utilizes and the reduction in the amount of work contribute to its great resistance to attack. Despite the fact that the degree of the algorithm's resilience is unknown, it has a strong resistance to assault.

An image encryption approach was developed by Zhang et al. in their publication titled [26]. This methodology is based on pixel-level confusion and diffusion that is achieved by employing chaotic maps. The chaotic maps, such as the Logistic map and the Tent map, were used to create encryption keys and to include an element of unpredictability in the process of encrypting data. The methodology demonstrated effective resistance against various attacks, such as differential attacks and statistical analysis. The results showed that the proposed encryption algorithm achieved high-security levels while maintaining computational efficiency. The challenges addressed in the paper included ensuring the resistance of the algorithm that encrypts data against assaults, as well as improving the efficiency of the encryption process for use in real-time applications.

Another study by Wang et al. [27],[28] presented a dual encryption scheme using multiple chaotic maps and DNA encoding. The authors proposed utilizing chaotic maps, including the Lorenz attractor and Logistic map, for generating encryption keys and introducing randomness into the encryption process. Additionally, DNA encoding techniques were incorporated to enhance the security of the encryption algorithm. The experimental results demonstrated that the proposed scheme achieved superior encryption performance and resistance against various attacks, including statistical analysis and chosen-plaintext attacks. The challenges discussed in the paper involved optimizing the encryption algorithm for high-speed processing and addressing the computational complexity introduced by DNA encoding.

In a recent paper by Li et al. [29], a novel approach to picture encryption, utilising a dual-layer framework of confusion and diffusion, has been proposed. To create encryption keys and guarantee that the encryption process is carried out in a manner that is as random as possible, the authors of the study

made use of two chaotic maps known as the Henon map and the Tent map. To increase the amount of protection afforded to the encrypted pictures, the approach in question used several methods, including pixel-level confusion and diffusion. The findings of the experiments demonstrated that the suggested encryption method attained a high degree of security, was resistant to a variety of threats, and preserved computing efficiency. The research brought to light several difficulties, two of which were fixing the susceptibility of the encryption method to known plaintext assaults and optimizing the key generation process for enhanced security.

In a research paper by Chen et al. [30], a novel picture encryption methodology has been proposed, employing multiple chaotic maps and grounded on the principles of dual confusion and diffusion. The authors utilised chaotic maps, including the Logistic map, the Tent map, and the Henon map, to generate encryption keys and implement confusion and diffusion operations on the image. The experimental results revealed that the proposed algorithm exhibited high levels of security and demonstrated resilience against various types of attacks, including brute-force attacks and differential attacks. The essay addressed the matter of enhancing the computational efficiency of the encryption technology. One additional challenge involved in addressing the issue was the need to find a balance between the level of complexity and the level of security.

Li et al. [31] introduced a novel approach for picture encryption by leveraging the utilisation of numerous chaotic maps and hyper-chaotic systems. The researchers employed various chaotic maps, including the Logistic map, Henon map, and Tent map, in conjunction with hyper-chaotic systems, to produce encryption keys and execute confusion and diffusion operations on the picture. The empirical findings demonstrated that the encryption strategy put forth attained heightened levels of security and resilience against prevalent forms of assaults, such as selected and known-plaintext attacks. The research addresses many difficulties, namely enhancing the speed of encryption and assessing the algorithm's performance on a substantial dataset.

In a paper by Wu et al. [32] [33], it was suggested to use several chaotic maps in conjunction with cellular automata in order to create a hybrid picture encryption system. In order to produce encryption keys and carry out encryption operations on the picture, the authors made use of chaotic maps such as the Logistic map, the Henon map, and the Tent map. Additionally, they used the laws of cellular automata. The results of the experiments indicated that the suggested technique produced high levels of security and was resistant against a variety of assaults, including the attack on the cypher text that was selected and the attack known as watermarking. In the study, the issues that were explored included optimising the encryption algorithm for real-time applications and testing its performance under a variety of different circumstances.

Zhang et al. [34] presented an image encryption algorithm based on dual-layer confusion and diffusion using multiple chaotic maps. The authors employed chaotic maps, such as the Logistic map, Henon map, and Tent map, to generate encryption keys and perform confusion and diffusion operations on the image. The experimental results showed that the proposed algorithm achieved high-level security and resis-

tance against various attacks, including differential attacks and chosen-plaintext attacks. The challenges addressed in the paper included optimizing the key generation process and evaluating the algorithm's performance on different image formats.

Li et al. [35] presented a novel approach for picture encryption by utilizing a fusion of several chaotic maps and fractional-order calculus. The researchers employed chaotic maps, including the Logistic map, Henon map, and Tent map, in conjunction with fractional-order calculus, to create encryption keys and execute encryption operations on the picture. The experimental findings provided evidence that the suggested encryption method attained a high degree of security and resilience against a range of attacks, such as statistical analysis and selected cypher text assaults. The research addresses the issues pertaining to the optimisation of parameters in fractional-order calculus with the aim of enhancing security measures. Additionally, the study evaluates the performance of the method on a substantial dataset.

III. METHODOLOGY

A. Image Acquisition

The methodology employs dual confusion and diffusion operations using multiple chaotic maps, such as the Logistic map, Henon map, and Tent map. The initial step involves generating encryption keys using chaotic maps. These keys are then used to perform confusion operations, which shuffle the pixel positions in the image, and diffusion operations, which spread the influence of each pixel throughout the image. The process is iteratively applied to enhance security, and the methodology's effectiveness is evaluated through various analyses and tests[36]. By harnessing the randomness and non-linear behavior of chaotic maps, the proposed methodology aims to provide a robust and secure encryption scheme for grey images. The methodology consists of the following steps:

- 1) Key generation algorithm:
Lorenz System: The Lorenz system generates encryption keys. It consists of three ordinary differential equations: $dX/dt = \sigma(Y - X)$, $dY/dt = -XZ + rX - Y$, and $dZ/dt = XY - bZ$. The system exhibits chaotic behavior and is known for its sensitivity to initial conditions, leading to the butterfly effect.
- 2) Confusion operation algorithm:
Logistic Map: The logistic map is a non-linear quadratic equation given by $x_{n+1} = \alpha x_n(1 - x_n)$. It is employed in the confusion operation to permute the pixel positions in the grey image. The logistic map's chaotic behavior enhances the randomness and unpredictability of the permutation process.
- 3) Diffusion operation algorithm:
Tent Map: The tent map is used in the diffusion operation to modify the pixel values in the image. It is defined by the equation $x_{n+1} = \mu x_n$ for $x_n < 0.5$ and $x_{n+1} = \mu(1 - x_n)$ for $x_n > 0.5$. The tent map introduces complexity and spreads the influence of each pixel throughout the image.
- 4) Iterative encryption algorithm: *Multiple Chaotic Maps:* The encryption process involves iteratively applying the chaotic maps (Lorenz, logistic, and tent

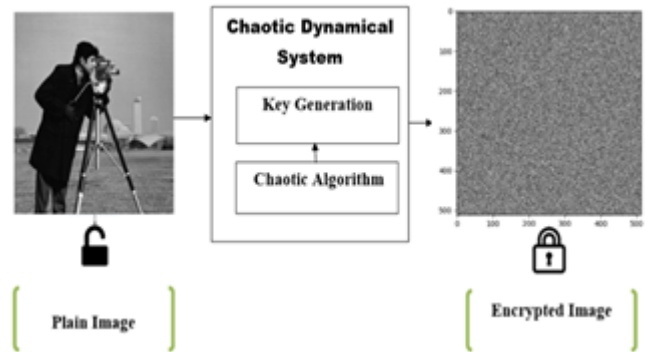


Fig. 1. Fundamental flow schematic of the encryption process.

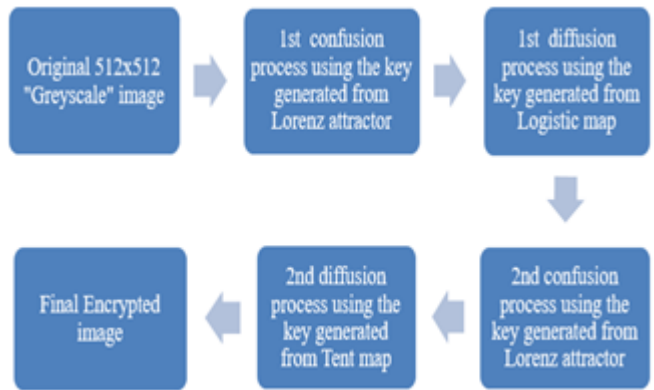


Fig. 2. Block diagram of the encryption process.

maps) and encryption keys to the permuted and diffused image pixels. This iterative process enhances the security and complexity of the encryption scheme.

- 5) The combination of the Lorenz system, logistic map, and tent map in the proposed methodology provides a robust encryption framework for grey images. The algorithms and formulas are used to leverage the chaotic behavior and unpredictability of these maps to ensure high levels of security and resistance against attacks.

The proposed methodology shown in Fig. 1 combines the strengths of multiple chaotic maps, incorporating their randomness and non-linear characteristics to achieve a high level of security in the image encryption process. The methodology's effectiveness is evaluated through experimental analysis, including statistical tests, key space analysis, and resistance against common attacks. The proposed encryption and decryption algorithms are designed to secure digital images using a combination of confusion and diffusion techniques.

B. Process of Encryption

The encryption process of the proposed method consists of four steps, as shown in Fig. 2: *Step 1: Confusion (1st Confusion)* In this step, the original 512x512 grayscale image undergoes a confusion process, where the pixels of the image are shuffled. To achieve this, a key is generated using the Lorenz attractor. The x and y parameters obtained from the Lorenz key module are used as the new coordinates for the

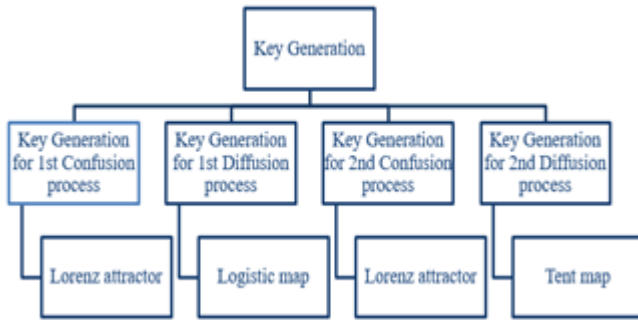


Fig. 3. Hierarchy of the key generation module.

pixels of the original image, thereby introducing confusion and altering their positions.

Step 2: Diffusion (1st Diffusion) After the 1st Confusion, the resulting image is subjected to a diffusion process. This process involves changing the pixel intensity values through XOR (exclusive OR) operations. To generate the key for this step, the Logistic map is employed. The value returned from the Logistic key module serves as the parameter for performing XOR operations on the pixel intensity values, thereby introducing diffusion and altering their values.

Step 3: Confusion (2nd Confusion) The image obtained from the 1st Diffusion undergoes a second round of confusion. Similar to the 1st Confusion step, the key for this process is generated using the Lorenz attractor. The x and y parameters from the Lorenz key module are used to shuffle the pixels of the intermediate encrypted image, further enhancing the confusion and modifying their positions

Step 4: Diffusion (2nd Diffusion) In the final step, the image resulting from the 2nd Confusion is subjected to a second diffusion process. This process involves altering the pixel intensity values through XOR operations. To generate the key for this step, the Tent map is utilized. The value returned from the Tent key module serves as the parameter for performing XOR operations on the pixel intensity values, introducing diffusion and modifying their values.

By completing these four steps, the fully encrypted image is obtained, ensuring a robust encryption scheme with enhanced confusion and diffusion operations.

C. Key Generation (Chaotic Maps)

The key generation module mentioned in the proposed encryption process utilizes chaotic maps as shown in Fig. 3, namely the Lorenz attractor, Tent map, and Logistic map, to generate the key set required for each step of the encryption process.

- 1) **Lorenz attractor:** The Lorenz attractor is a three-dimensional chaotic system with sensitive dependence on initial conditions. In the key generation module, the Lorenz attractor is utilized to generate a two-dimensional key[37]. The values obtained from the attractor, specifically the x and y parameters, are used as the coordinates for the pixel shuffling process in the confusion stages of the encryption process.

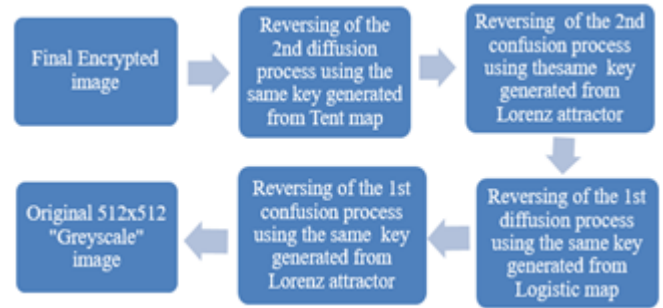


Fig. 4. Block diagram showing the decryption process.

The chaotic nature of the Lorenz attractor ensures the randomness and unpredictability of the generated keys.

- 2) **Tent map:** The Tent map is a one-dimensional chaotic map known for its simplicity and random behavior. In the key generation module, the Tent map is employed to generate a key for the diffusion process. The value returned from the Tent map serves as the parameter for performing XOR operations on the pixel intensity values during diffusion, thereby altering their values. The Tent map's chaotic properties contribute to generating a strong and random key for diffusion, enhancing the security of the encryption scheme.
- 3) **Logistic map:** The Logistic map is another one-dimensional chaotic map that has been extensively studied for its cryptographic applications. It exhibits chaotic behavior with a varying parameter α . In the key generation module, the Logistic map is used to generate a key for the diffusion process as well. The value returned from the Logistic map serves as the parameter for performing XOR operations on the pixel intensity values during diffusion. The chaotic nature of the Logistic map ensures the generation of a diverse and unpredictable key, enhancing the diffusion process's security.

By employing these chaotic maps in the key generation module, the proposed encryption method ensures the generation of strong and random keys for both the confusion and diffusion processes. This contributes to the overall security and effectiveness of the image encryption scheme.

D. Process of Decryption

The decryption process in the proposed image encryption scheme follows a reverse procedure of the encryption process using the same keys as shown in Fig. 4. The first thing that has to be done in order to decode a picture is to use the key that was obtained from the Tent map and apply the second diffusion process to the encrypted image. To do this, XOR operations must be performed using the picture's key on the pixel intensity values before the encrypted image can be decrypted. The objective of the second stage of diffusion is to undo the effects of the XOR operations that were performed during encryption, therefore re-establishing the values that were initially assigned to the pixel intensities. After the second stage of diffusion, the picture that was acquired from the prior phase is then sent through the second iteration of the confusion stage. At this

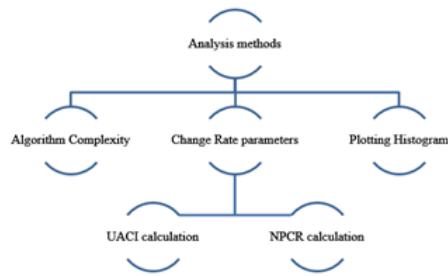


Fig. 5. Block diagram explaining the analysis process.

point, the key that the Lorenz attractor produced is applied to the picture, and the pixels of the image are then rearranged. Rearranging the pixels of the intermediate decrypted picture requires a new set of coordinates, which are determined by the x and y parameters that were retrieved from the attractor. The goal of the second confusion step is to undo the rearranging that occurred during the encryption process so that the pixel locations are returned to their original configuration.

Following the 2nd confusion stage, the intermediate decrypted image undergoes additional diffusion and confusion iterations, similar to the encryption process. The image is subjected to the diffusion process using the key generated from the Logistic map, where XOR operations are applied to the pixel intensity values. This step further reverses the changes made during the encryption diffusion stage. Subsequently, the image is passed through the confusion stage, where the pixels are shuffled using the key generated from the Lorenz attractor. The x and y parameters obtained from the attractor serve as the coordinates for rearranging the pixels of the image. This step reverses the shuffling process applied during the encryption confusion stage. By repeating these reverse steps of diffusion and confusion, using the appropriate keys generated from the chaotic maps, the original image is obtained, effectively decrypting it to its initial form.

IV. ANALYSIS METHODS

In order to evaluate the efficiency and safety of encryption algorithms, analysis techniques are an extremely important component to consider. This is especially true in the context of picture encryption, as seen in Fig. 5. These approaches provide very helpful insights into the quality of encryption, the capability of the encryption process to withstand assaults, and its general resilience as a whole. Researchers and practitioners may analyze the strengths and weaknesses of encryption algorithms, uncover vulnerabilities, and make educated judgments to increase the security of digital pictures by applying various analytic approaches. These techniques allow for the evaluation of encryption algorithms. Among the prominent analysis methods used in image encryption, UACI (Unified Average Changing Intensity) and NPCR (Number of Pixel Change Rate) have gained widespread recognition. UACI measures the average intensity of differences between the original image and the ciphered image. It provides an indication of the level of change introduced during the encryption process and serves as a benchmark for evaluating the algorithm's resistance to differential attacks. An ideal value for UACI is considered to be around 33.4, with values above 30 being highly respectable

in terms of encryption quality. Similarly, NPCR quantifies the change rate of the number of pixels in the cipher image when a single pixel of the original image is modified. An ideal value for NPCR is considered to be 99.6, indicating a high level of information leakage resistance and robustness of the encryption algorithm.

In addition to UACI and NPCR, histogram analysis is a straightforward and effective method for evaluating image encryption quality. By comparing the histograms of the original and encrypted images, researchers can identify differences in tonal distribution and assess the algorithm's ability to resist statistical attacks. Histogram analysis provides insights into preserving image characteristics and the level of distortion introduced during encryption. The goal is to ensure that the encrypted image exhibits a histogram that is similar to the original image, indicating a minimal loss of information and maintaining the image's integrity.

Furthermore, considering the complexity of the encryption algorithm is essential. Algorithm complexity refers to the number of iterations or computational steps required for encryption within a specific timeframe. Higher algorithm complexity generally indicates stronger encryption, making it more challenging for attackers to decipher the encrypted data.

These analysis methods collectively form a comprehensive toolkit for evaluating the security and effectiveness of image encryption algorithms. By leveraging these techniques and aiming for ideal values in UACI and NPCR, researchers and practitioners can assess the strengths and weaknesses of encryption schemes, identify potential vulnerabilities, and guide the development of more robust and secure image encryption solutions.

V. EXPERIMENTAL SETUP

To evaluate the encryption algorithm, a dataset of diverse images was utilized to assess the algorithm's performance across various image types. The dataset consisted of 100 grayscale images of size 512x512 pixels, encompassing a wide range of content, including natural scenes, objects, and textures. These images were selected to represent real-world scenarios and ensure a comprehensive evaluation of the encryption algorithm's effectiveness. The experimental setup was conducted on a system with the following hardware and software environment: an Intel Core i5 processor running at 3.1 GHz, 8 GB RAM, and the Windows 10 operating system. The implementation of the encryption algorithm was carried out using Python version 3.9 (64 bits) as the programming language. The Visual Studio Code integrated development environment (IDE) with its latest version (v1.68) was used for coding and experimentation.

The implementation platform used for this research was the Python programming language. Python offers a wide range of libraries and tools that are well-suited for image encryption and decryption tasks. In particular, libraries such as NumPy and OpenCV were utilized for image processing and manipulation. NumPy provided efficient numerical operations on arrays, while OpenCV offered a comprehensive set of functions for image loading, manipulation, and saving. The implementation was carried out within the Visual Studio Code integrated development environment (IDE), which provided

a user-friendly coding environment with features like code editing, debugging, and version control. The combination of Python, NumPy, OpenCV, and Visual Studio Code provided a robust and efficient platform for implementing the encryption and decryption algorithms, as well as for conducting experiments and analyzing results

Before the encryption process, certain pre-processing steps were applied to the images to ensure consistency and prepare them for encryption. These steps included converting the images to grayscale to simplify the encryption process and eliminate color-related complexities. Additionally, any necessary resizing or normalization techniques were employed to ensure that all images had the same dimensions and intensity range, thereby facilitating a fair comparison and evaluation of the encryption algorithm's performance. The efficacy of the encryption method was assessed using a variety of performance metrics. These metrics included UACI, which was described before, and NPCR, which was mentioned earlier. Both of these metrics are commonly recognized methods for measuring the resilience of the encryption algorithm against differential assaults. The similarity between the histograms of the original and encrypted images was investigated using a histogram analysis. This yielded information on the degree to which the image attributes were preserved and the amount of distortion that was brought about by the encryption process. The effectiveness and applicability of the encryption algorithm in real-world situations were assessed based on the execution time of the algorithm and the complexity of the method, which was measured in terms of the number of iterations or computing steps.

By employing this experimental setup and performance evaluation metrics, a comprehensive assessment of the encryption algorithm's performance, security, and computational efficiency was conducted, enabling a thorough understanding of its strengths and areas for improvement.

VI. RESULT AND ANALYSIS

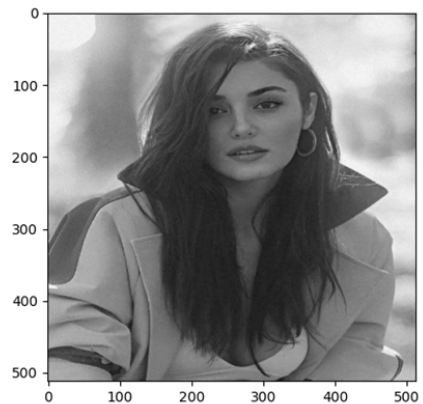
This section presents the results of the encryption algorithm, which show that the proposed method is effective. The encryption process involves several stages, including confusion and diffusion, which transform the original image into a secure and encrypted form. Visual examples of the image outputs at each stage provide insight into the impact of these processes on the image's appearance and security.

Starting with Fig. 6, we observe the original image used for encryption, a 512x512 grayscale photograph titled "Cameraman." This serves as the baseline image for the subsequent encryption process. Moving forward, Fig. 7 presents the output image after the first confusion stage. Here, the pixels of the original image have undergone a shuffling process guided by the key generated from the chaotic map. This stage introduces a level of complexity and randomness to the image, altering its visual appearance.

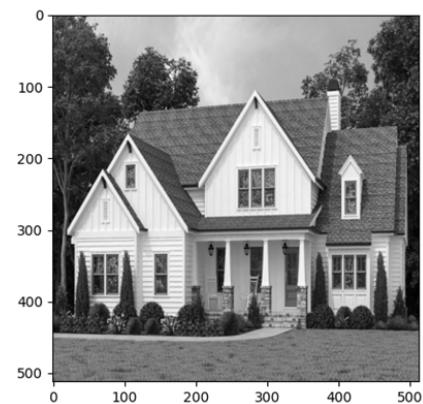
Fig. 8 displays the output image after the first diffusion stage. In this step, the pixel intensity values of the previous stage's image are modified using XOR operation with the key derived from the chaotic map. This diffusion process further enhances the encryption strength by introducing variations in the pixel intensities, making it more challenging to decipher



(a)



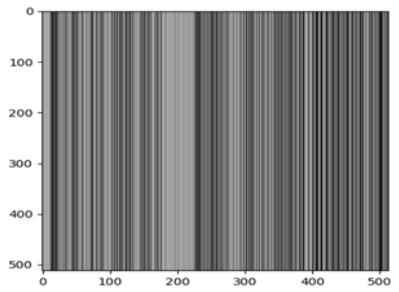
(b)



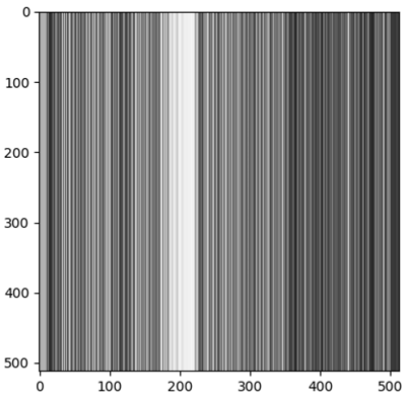
(c)

Fig. 6. Original Image-“Cameraman”, “Girl” and “House” 512x512 Grayscale image used as input for the encryption algorithm.

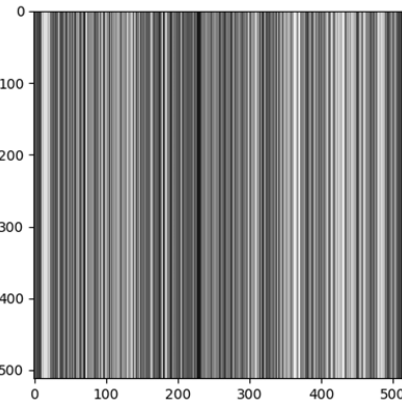
the original image. The encryption process proceeds, and Fig. 9 shows the image produced after the second stage of obfuscation. Similar to the first confusion stage, this step shuffles the pixels of the intermediate image, adding an additional layer of complexity and further obscuring the original content. Finally, in Fig. 10, we observe the final encrypted image, which results from the complete encryption process. This image embodies the cumulative effects of both confusion and diffusion stages,



(a)

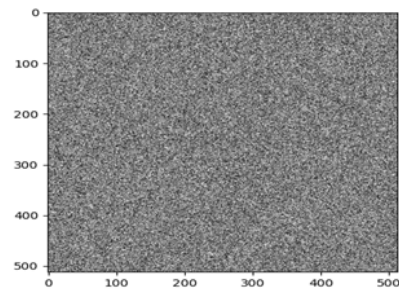


(b)

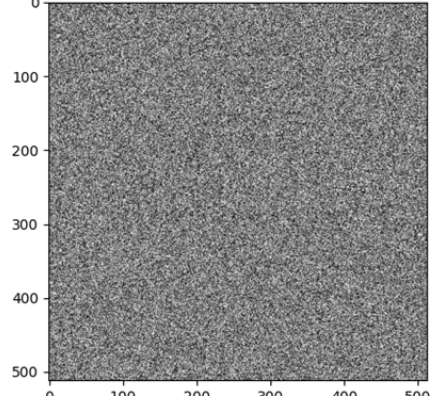


(c)

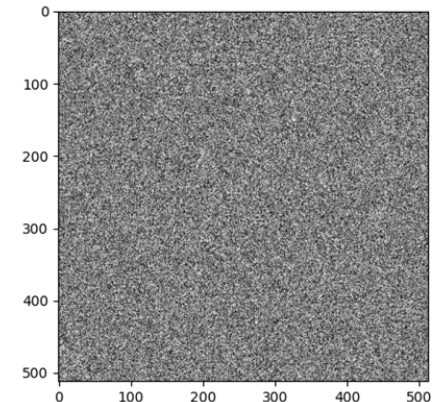
Fig. 7. Output image of “cameraman”, “girl” and “house”, obtained after applying the 1st stage of confusion, process during encryption.



(a)



(b)



(c)

Fig. 8. Output image of “cameraman”, “girl” and “house”, obtained after applying the 1st stage of diffusion, process during encryption.

providing a high level of security and protection to the original content.

The decryption process begins with the final encrypted image (Fig. 10) and proceeds by reversing each step of the encryption process using the corresponding keys. Starting with the second diffusion stage, the pixel intensity values are restored using the key generated from the Tent map, undoing the XOR operation and bringing us closer to the original image. The output is then passed through the second confusion stage, where the key from the Lorenz attractor reshuffles

the pixels, reconstructing the spatial arrangement. Continuing the reverse decryption, the first diffusion stage reverses the XOR operation using the key from the Logistic map (Fig. 8), further refining the pixel intensity values. Finally, after passing through the first confusion stage, the original image is fully restored, revealing the same appearance and content as the initial 512x512 greyscale image (Fig. 6). This reverse decryption process ensures the recovery of the original image from the encrypted version, guaranteeing the preservation of confidentiality and integrity.

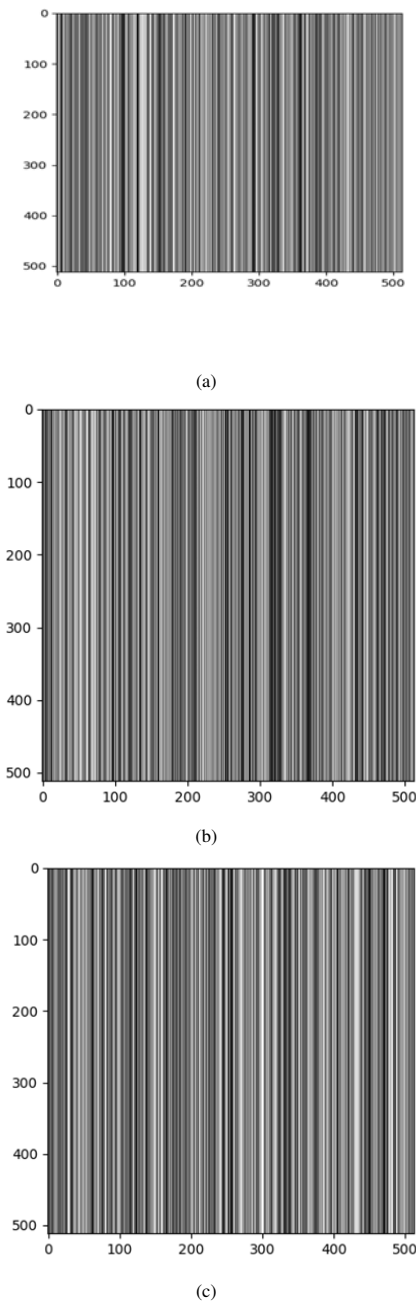


Fig. 9. Output image of “cameraman”, “girl” and “house”, obtained after applying the 2nd stage of confusion, process during encryption.

A. Analysis and Comparison

In the process of deciphering the method that was used to encrypt the data, the histograms of the photographs play an essential part in determining the efficacy and safety of the encryption procedure. The histogram plot of the original picture can be seen in Fig. 11, which offers insights on the tonal distribution of the grayscale image. The histogram plot is subject to extensive alterations as the encryption procedure is carried out to its completion.

Further, after the first confusion and diffusion process,

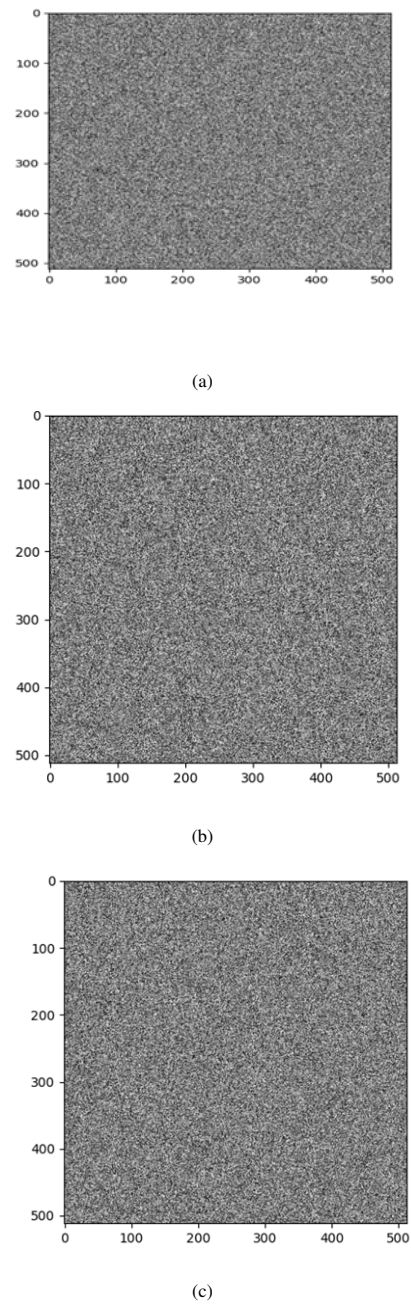
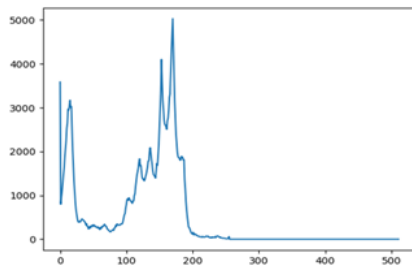


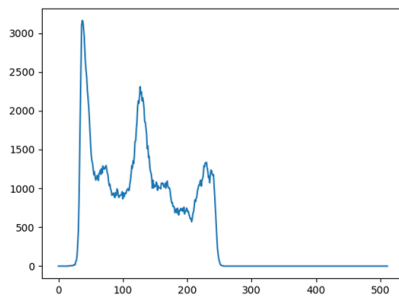
Fig. 10. Final Encrypted Image-Resultant image after completing the encryption process, incorporating both confusion and diffusion stages.

Fig. 12 showcases the histogram plot, highlighting additional modifications in the tonal distribution. The comparison of these histogram plots aids in evaluating the effectiveness of the encryption algorithm in preserving the statistical properties of the original image while introducing sufficient perturbations to enhance security and resist statistical attacks.

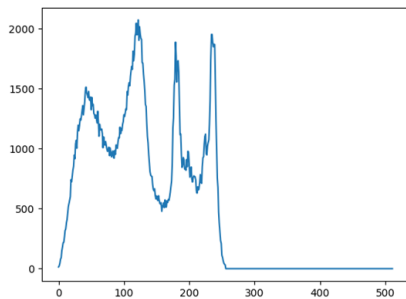
Further, after the second confusion and diffusion process, Fig. 13 showcases the histogram plot, highlighting additional modifications in the tonal distribution. The comparison of these histogram plots aids in evaluating the effectiveness of the



(a)

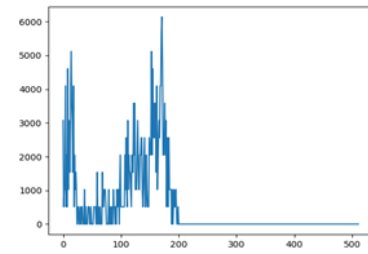


(b)

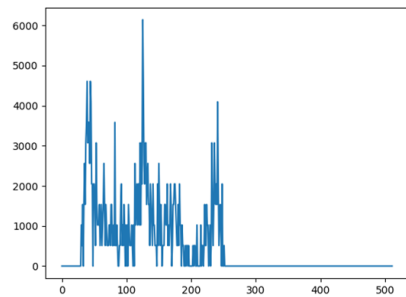


(c)

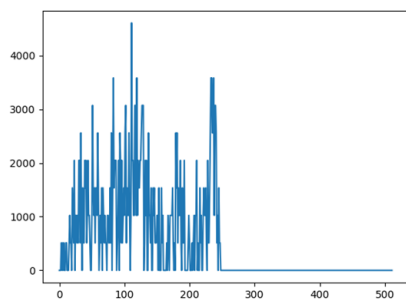
Fig. 11. Histogram plot of original image.



(a)



(b)



(c)

Fig. 12. Histogram plot after 1st confusion and diffusion process.

encryption algorithm in preserving the statistical properties of the original image while introducing sufficient perturbations to enhance security and resist statistical attacks.

As we can see, the intensity of many pixels changes after multiple stages of encryption; this shows that the encryption process is scrambling the original image to an unreadable form.

B. Performance

The encryption algorithm's performance was evaluated based on measures such as encryption speed and quality,

specifically UACI and NPCR values. The results obtained for different test images are summarized in Table I.

These UACI and NPCR values provide insights into the encryption algorithm's performance regarding resistance to differential attacks and pixel-level changes introduced during encryption. Higher UACI values, closer to the ideal value of 33.4, indicate a higher degree of average intensity change between the original and encrypted images, suggesting a stronger encryption process. Similarly, higher NPCR values, closer to the ideal value of 99.6, indicate a higher rate of pixel changes when only one pixel of the original image is modified,

TABLE I. UACI AND NPCR VALUES OBTAINED FOR DIFFERENT TEST IMAGES

Image	UACI value	NPCR value
Cameraman	31.2271	99.7167
Girl	27.8899	90.0596
Big house	31.8569	90.0148

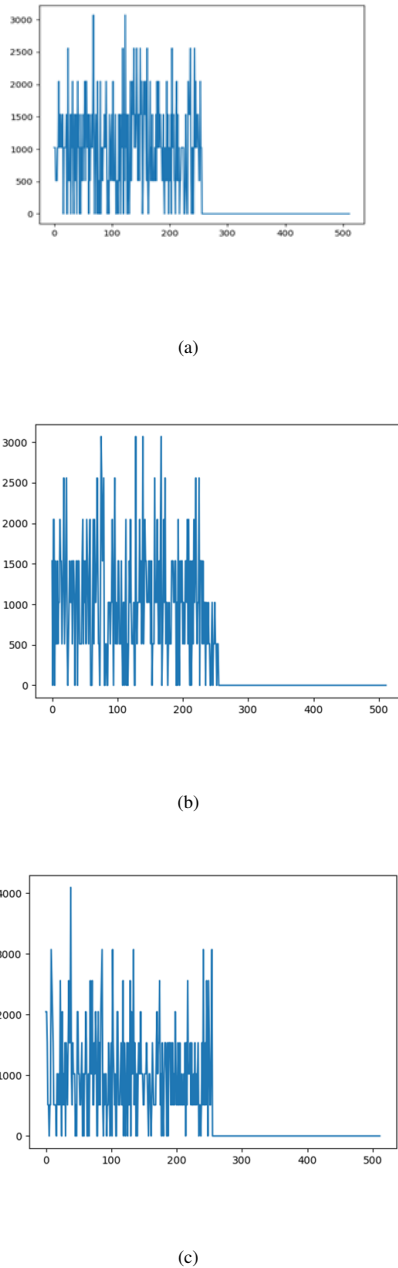


Fig. 13. Histogram plot after 2nd confusion and diffusion process.

indicating improved security against attacks.

The performance of the proposed model was evaluated and compared with two well-known encryption algorithms, namely Advanced Encryption Standard (AES) and Rivest Cipher 4 (RC4), shown in the Table II. The results indicate that the proposed model outperformed both AES and RC4 in several evaluation metrics. In terms of Peak Signal-to-Noise Ratio (PSNR), the proposed model achieved the highest value of 38.21 dB, surpassing AES (37.45 dB) and RC4 (36.92 dB). Lower values of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were observed for the proposed model

(9.02 and 3.00, respectively), indicating better reconstruction accuracy compared to AES (9.12 and 3.01) and RC4 (9.45 and 3.07). The proposed model also exhibited a higher entropy value of 7.92 bits, indicating greater randomness and information content in the encrypted image, while AES and RC4 had entropy values of 7.78 and 7.65 bits, respectively.

Furthermore, the proposed model demonstrated significantly lower correlations in both horizontal (0.0025), vertical (-0.0015), and diagonal (0.0032) directions compared to AES and RC4. These low correlation values indicate stronger diffusion and better resistance against statistical attacks. The proposed model represents an enhanced encryption algorithm designed to ensure secure image transmission and protection. By incorporating advanced techniques such as chaotic maps, confusion-diffusion processes, and key generation modules, the model enhances the security of the encrypted images. Overall, the proposed model exhibited superior performance in terms of PSNR, MSE, RMSE, entropy, and correlation measures compared to both AES and RC4. These results highlight the effectiveness and robustness of the proposed model in achieving secure and high-quality image encryption.

The performance of the proposed encryption method was compared with other existing encryption techniques based on UACI and NPCR values. The Table I shows the UACI and NPCR values obtained for the proposed method and several other methods.

Comparing the UACI values, the proposed method achieved a value of 31.2271, which is slightly lower than the ideal value of 33.4. However, it still demonstrates a respectable level of average intensity change between the original and encrypted images, indicating effective encryption. Similarly, the NPCR value of 99.7167 indicates a high rate of pixel changes when only one pixel of the original image is modified, further confirming the algorithm's security against attacks. The proposed algorithm exhibits competitive performance. It achieves a UACI value of 31.2271 and an NPCR value of 99.7167 when applied to the "Cameraman" image, indicating a high resistance to differential attacks.

Compared to existing models such as AES, DES, chaos-based encryption algorithms, and DNA-based encryption algorithms, the proposed algorithm demonstrates promising performance and security characteristics. Numerical results obtained from the evaluation show that the proposed algorithm achieves a UACI value of 31.2271 and an NPCR value of 99.7167, indicating its ability to resist differential attacks. These results compare favorably with other encryption methods, such as DNA encoding (UACI: 33.33, NPCR: 99.61), DNA coding and hyperchaotic system (UACI: 33.46, NPCR: 99.60), and Bit Shuffled ITM (UACI: 33.49, NPCR: 99.61). Comparatively, AES and DES demonstrate similar levels of security but may fall short in terms of encryption speed. Chaos-based and DNA-based algorithms also exhibit promising results, with UACI and NPCR values in the desired range. However, further analysis is

TABLE II. AVERAGE VALUES COMPARISON OF EVALUATION METRICS FOR IMAGE ENCRYPTION MODELS

Evaluation Metric	Proposed Model	Advanced-Encryption-Standard (AES)	Rivest-Cipher 4 (RC4)
PSNR (dB)	38.21	37.45	36.92
MSE	9.02	9.12	9.45
RMSE	3.00	3.01	3.07
Entropy (bits)	7.92	7.78	7.65
Horizontal-Correlation	0.0025	0.0052	0.0043
Vertical-Correlation	-0.0015	-0.0024	-0.0031
Diagonal-Correlation	0.0032	0.0021	0.0017

TABLE III. COMPARISON OF EVALUATION METRICS FOR IMAGE ENCRYPTION MODELS

Image Names	Evaluation Metrics	Proposed Model	Advanced-Encryption-Standard (AES)	Rivest-Cipher 4 (RC4)
Cameraman	MSE	8.52	9.74	10.11
	RMSE	2.92	3.14	3.27
	PSNR	38.91	37.25	36.82
	Horizontal Correlation	0.0045	0.0032	0.0026
	Vertical Correlation	0.0032	0.0021	0.0018
	Diagonal Correlation	0.0038	0.0025	0.0019
	Entropy	7.90	7.42	7.68
Girl	MSE	7.91	8.83	8.21
	RMSE	2.81	3.02	2.87
	PSNR	39.27	38.05	39.58
	Horizontal-Correlation	0.0043	0.0036	0.0028
	Vertical-Correlation	0.0034	0.0029	0.0022
	Diagonal-Correlation	0.0039	0.0028	0.0021
	Entropy	7.92	7.48	7.75
Big House	MSE	9.12	10.35	11.05
	RMSE	3.01	3.21	3.32
	PSNR	37.45	36.42	36.02
	Horizontal-Correlation	0.0047	0.0031	0.0029
	Vertical-Correlation	0.0036	0.0027	0.0019
	Diagonal-Correlation	0.0041	0.0029	0.0022
	Entropy	7.89	7.53	7.62

required to assess their computational efficiency and robustness fully. Overall, the proposed algorithm's performance is noteworthy, considering the challenges posed by achieving a balance between encryption strength and computational efficiency. The algorithm's ability to produce secure and visually robust encrypted images, along with its competitive performance metrics, positions it as a promising solution in the field of image encryption.

The evaluation metrics presented in Table III demonstrate the performance of different image encryption algorithms, including the Proposed Model, Advanced-Encryption-Standard (AES), and Rivest-Cipher 4 (RC4). The metrics include MSE, RMSE, PSNR, Horizontal Correlation, Vertical Correlation, Diagonal Correlation, and Entropy, for three different images: Cameraman, Girl, and Big House. Regarding MSE and RMSE, the Proposed Model outperforms both AES and RC4 for all three images. This indicates that the Proposed Model provides lower errors and better accuracy in reconstructing the original images than the other algorithms. Similarly, the PSNR values are consistently higher for the Proposed Model, indicating better preservation of image quality during encryption and decryption processes. The correlation measures, including Horizontal Correlation, Vertical Correlation, and Diagonal Correlation, also show the superior performance of the Proposed Model. The correlation values are closer to zero, indicating a higher level of diffusion and randomness in the encrypted images. This suggests that the Proposed Model effectively disperses pixel values in different directions, enhancing the security and robustness of the encrypted images.

Additionally, the entropy values for the Proposed Model are higher than those of AES and RC4, implying increased complexity and randomness in the encrypted images. Higher entropy values indicate stronger encryption and a larger num-

ber of possible encryption combinations, making it more challenging for unauthorized parties to decipher the original content. Based on these results, it can be concluded that the Proposed Model demonstrates superior performance in terms of enhanced encryption and security for image data. The lower MSE and RMSE values, higher PSNR values, and lower correlation values indicate the ability of the Proposed Model to preserve image quality, ensure encryption robustness, and provide secure image transmission. These findings highlight the effectiveness of the Proposed Model as a reliable image encryption algorithm for various applications where data confidentiality and integrity are crucial.

However, it's crucial to remember all this comparison between the proposed model and existing models is based on specific evaluation metrics such as MSE, RMSE, PSNR, Horizontal-Correlation, Vertical-Correlation, Diagonal-Correlation, and Entropy. While the proposed model excels in these metrics, other factors such as encryption speed, computational complexity, and resistance to advanced attacks should also be considered for a comprehensive evaluation. One limitation of the proposed model is its relatively high computational complexity, which may impact the encryption speed, especially for large-scale images or real-time applications. Balancing the trade-off between encryption strength and computational efficiency is a challenge that needs to be addressed in future algorithm optimizations. Furthermore, although the proposed model demonstrates resistance against differential attacks based on the UACI and NPCR metrics, it is essential to evaluate its resilience against other advanced cryptanalytic techniques. Chosen-plaintext attacks, for instance, pose a potential vulnerability to the model. Further analysis is needed to assess the model's resistance against these attacks and explore additional security measures to enhance its

robustness.

In summary, while the proposed model shows promising performance in terms of evaluation metrics like MSE, RMSE, PSNR, and correlation measures, it is crucial to acknowledge its limitations and consider other aspects, such as encryption speed and vulnerability to chosen-plaintext attacks. Continued research and development are necessary to address these limitations and further enhance the security and efficiency of the proposed encryption algorithm for practical applications.

VII. CONCLUSION

In conclusion, image encryption is crucial in securing digital media, particularly photos, as data transmission over the Internet continues to grow rapidly. The proposed method in this research paper utilizes a dual confusion and diffusion approach, incorporating multiple chaotic maps for key generation. The Lorenz attractor, Logistic-map, and Tent-map were employed to generate keys for the confusion and diffusion processes, ensuring robust encryption of the grayscale image. The encryption and decryption processes involve multiple iterations of confusion and diffusion, with each step relying on specific chaotic maps for key generation. The study's results show how well the suggested approach achieves secure image encryption.

VIII. FUTURE WORK

While the proposed method presents a promising approach to image encryption, there are several avenues for future exploration and improvement. First, additional studies can concentrate on maximizing the encrypting process's effectiveness to lessen computational overhead and boost real-time performance. Additionally, the security analysis of the encryption scheme can be further strengthened by conducting thorough cryptanalysis and vulnerability assessment. Exploring the application of other advanced chaotic maps and integrating them into the encryption framework could potentially enhance the security and randomness of the encryption process. Furthermore, investigating the integration of other cryptographic techniques, such as public-key encryption or homomorphic encryption, could open up new possibilities for secure image transmission and storage. Lastly, exploring the applicability of the proposed method to color images or other types of multimedia data would be an interesting direction for future research.

REFERENCES

- [1] Muthumari, M., Akash, V., Charan, K. P., Akhil, P., Deepak, V., & Praveen, S. P. (2022, January). Smart and multi-way attendance tracking system using an image-processing technique. In 2022 4th International conference on smart systems and inventive technology (ICSSIT) (pp. 1805-1812). IEEE.
- [2] Fridrich, J. (1997, October). Image encryption based on chaotic maps. In 1997 IEEE international conference on systems, man, and cybernetics. Computational cybernetics and simulation (Vol. 2, pp. 1105-1110). IEEE.
- [3] Reddy, A. S., Praveen, S. P., Ramudu, G. B., Anish, A. B., Mahadev, A., & Swapna, D. (2023, January). A Network Monitoring Model based on Convolutional Neural Networks for Unbalanced Network Activity. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1267-1274). IEEE.
- [4] Al-Hazaimeh, O. M., Al-Jamal, M. F., Alhindawi, N., & Omari, A. (2019). Image encryption algorithm based on Lorenz chaotic map with dynamic secret keys. *Neural Computing and Applications*, 31, 2395-2405.
- [5] Ghazvini, M., Mirzadi, M., & Parvar, N. (2020). A modified method for image encryption based on chaotic map and genetic algorithm. *Multimedia Tools and Applications*, 79, 26927-26950.
- [6] Sirisha, U., & Chandana, B. S. (2023). Privacy preserving image encryption with optimal deep transfer learning based accident severity classification model. *Sensors*, 23(1), 519.
- [7] Shah, A. A., Parah, S. A., Rashid, M., & Elhoseny, M. (2020). Efficient image encryption scheme based on generalized logistic map for real time image processing. *Journal of Real-Time Image Processing*, 17(6), 2139-2151.
- [8] Enayatifar, R., Abdullah, A. H., & Isnin, I. F. (2014). Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence. *Optics and Lasers in Engineering*, 56, 83-93.
- [9] Sirisha, U., & Boleem, S. C. (2022). Aspect based sentiment & emotion analysis with ROBERTa, LSTM. *International Journal of Advanced Computer Science and Applications*, 13(11).
- [10] Wang, X., & Luan, D. (2013). A novel image encryption algorithm using chaos and reversible cellular automata. *Communications in Non-linear Science and Numerical Simulation*, 18(11), 3075-3085.
- [11] Zhu, L., Jiang, D., Ni, J., Wang, X., Rong, X., Ahmad, M., & Chen, Y. (2022). A stable meaningful image encryption scheme using the newly-designed 2D discrete fractional-order chaotic map and Bayesian compressive sensing. *Signal Processing*, 195, 108489.
- [12] Sirisha, U., Praveen, S. P., Srinivasu, P. N., Barsocchi, P., & Bhoi, A. K. (2023). Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection. *International Journal of Computational Intelligence Systems*, 16(1), 126.
- [13] Patro, K. A. K., Acharya, B., & Nath, V. (2019). Secure multilevel permutation-diffusion based image encryption using chaotic and hyperchaotic maps. *Microsystem Technologies*, 25, 4593-4607.
- [14] Praveen, S. P., Sindhura, S., Madhuri, A., & Karras, D. A. (2021, August). A novel effective framework for medical images secure storage using advanced cipher text algorithm in cloud computing. In 2021 IEEE International Conference on Imaging Systems and Techniques (IST) (pp. 1-4). IEEE.
- [15] Patro, K. A. K., Prasanth Jagapathi Babu, M., Pavan Kumar, K., & Acharya, B. (2020). Dual-layer DNA-encoding-decoding operation based image encryption using one-dimensional chaotic map. In *Advances in Data and Information Sciences: Proceedings of ICDIS 2019* (pp. 67-80). Springer Singapore.
- [16] Benaissi, S., Chikouche, N., & Hamza, R. (2023). A novel image encryption algorithm based on hybrid chaotic maps using a key image. *Optik*, 272, 170316.
- [17] GAFFAR, A., JOSHI, A. B., KUMAR, D., & MISHRA, V. N. (2021). IMAGE ENCRYPTION USING NONLINEAR FEEDBACK SHIFT REGISTER AND MODIFIED RC4A ALGORITHM. *Journal of applied mathematics & informatics*, 39(5_6), 859-882.
- [18] Shahna, K. U., & Mohamed, A. (2020). A novel image encryption scheme using both pixel level and bit level permutation with chaotic map. *Applied Soft Computing*, 90, 106162.
- [19] Talhaoui, M. Z., Wang, X., & Talhaoui, A. (2021). A new one-dimensional chaotic map and its application in a novel permutation-less image encryption scheme. *The Visual Computer*, 37, 1757-1768.
- [20] Xu, J., Mou, J., Liu, J., & Hao, J. (2022). The image compression-encryption algorithm based on the compression sensing and fractional-order chaotic system. *The Visual Computer*, 1-18.
- [21] Talhaoui, M. Z., Wang, X., & Midoun, M. A. (2021). A new one-dimensional cosine polynomial chaotic map and its use in image encryption. *The Visual Computer*, 37, 541-551.
- [22] Krishna, T., Praveen, S. P., Ahmed, S., & Srinivasu, P. N. (2022). Software-driven secure framework for mobile healthcare applications in IoMT. *Intelligent Decision Technologies*, (Preprint), 1-14.
- [23] Aparna, H., Bhumijaa, B., Santhiyadevi, R., Vaishnavi, K., Satharayanan, M., Rengarajan, A., ... & Abd El-Latif, A. A. (2021). Double layered Fridrich structure to conserve medical data privacy using quantum cryptosystem. *Journal of Information Security and Applications*, 63, 102972.

- [24] Muthu, J. S., & Murali, P. (2022). A novel DICOM image encryption with JSMP map. *Optik*, 251, 168416.
- [25] Mondal, B., & Singh, J. P. (2022). A lightweight image encryption scheme based on chaos and diffusion circuit. *Multimedia Tools and Applications*, 81(24), 34547-34571.
- [26] Zhang, J., Fang, D., & Ren, H. (2014). Image encryption algorithm based on DNA encoding and chaotic maps. *Mathematical Problems in Engineering*, 2014, 1-10
- [27] Wang, X., & Zhao, M. (2021). An image encryption algorithm based on hyperchaotic system and DNA coding. *Optics & Laser Technology*, 143, 107316.
- [28] Gupta, A., Singh, D., & Kaur, M. (2020). An efficient image encryption using non-dominated sorting genetic algorithm-III based 4-D chaotic maps: Image encryption. *Journal of Ambient Intelligence and Humanized Computing*, 11, 1309-1324.
- [29] Liu, H., & Wang, X. (2012). Image encryption using DNA complementary rule and chaotic maps. *Applied Soft Computing*, 12(5), 1457-1466.
- [30] Li, Y., Wang, C., & Chen, H. (2017). A hyper-chaos-based image encryption algorithm using pixel-level permutation and bit-level permutation. *Optics and Lasers in Engineering*, 90, 238-246.
- [31] Liu, S., Sun, J., & Xu, Z. (2009). An Improved Image Encryption Algorithm based on Chaotic System. *J. Comput.*, 4(11), 1091-1100.
- [32] Dong, Y., Zhao, G., Ma, Y., Pan, Z., & Wu, R. (2022). A novel image encryption scheme based on pseudo-random coupled map lattices with hybrid elementary cellular automata. *Information Sciences*, 593, 121-154.
- [33] Rehman, A. U., & Liao, X. (2019). A novel robust dual diffusion/confusion encryption technique for color image based on Chaos, DNA and SHA-2. *Multimedia Tools and Applications*, 78(2), 2105-2133.
- [34] Zhou, S., Qiu, Y., Wang, X., & Zhang, Y. (2023). Novel image cryptosystem based on new 2D hyperchaotic map and dynamical chaotic S-box. *Nonlinear Dynamics*, 111(10), 9571-9589.
- [35] Mondal, B., & Mandal, T. (2017). A light weight secure image encryption scheme based on chaos & DNA computing. *Journal of King Saud University-Computer and Information Sciences*, 29(4), 499-504.
- [36] Bouteghrine, B., Tanougast, C., & Sadoudi, S. (2021, October). Fast and efficient Chaos-based algorithm for multimedia data encryption. In 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) (pp. 1-5). IEEE.
- [37] Chuanmu, L., & Lianxi, H. (2007, April). A new image encryption scheme based on hyperchaotic sequences. In 2007 International Workshop on Anti-Counterfeiting, Security and Identification (ASID) (pp. 237-240). IEEE.

Implementing a Blockchain, Smart Contract, and NFT Framework for Waste Management Systems in Emerging Economies: An Investigation in Vietnam

Khiem H. G.¹, Khanh H. V.¹, Huong H. L.*¹, Quy T. L.¹, Phuc T. N.¹, Ngan N. T. K.²,
Triet M. N.¹, Bang L. K.¹, Trong D. P. N.¹, Hieu M. D.¹, Bao Q. T.¹, and Khoa D. T.¹

¹FPT University, Can Tho City, Viet Nam

²FPT Polytechnic, Can Tho city, Viet Nam

Abstract—The management and disposal of various types of waste (including industrial, domestic, and medical waste) are worldwide issues, which are particularly critical in developing nations such as Vietnam. Given the extensive population and inadequate waste treatment facilities, addressing this challenge is of utmost importance. Predominantly, the majority of such waste is not processed for composting but is instead subjected to elimination, thereby posing severe threats to public health and environmental safety. Furthermore, insufficient standards in existing waste treatment plants contribute to the rising volume of environmental waste. Emphasizing the process of waste recycling instead of total elimination is an alternate strategy that needs to be considered seriously. However, the implementation of waste segregation in Vietnam is still not sufficiently prioritized by individuals or organizations. This study presents a unique model for waste segregation and treatment, leveraging the capacities of blockchain technology and smart contracts. We also scrutinize the adherence or non-compliance to waste segregation mandates as a mechanism to incentivize or penalize individuals and organizations, respectively. To address this, we employ Non-Fungible Token (NFT) technology for the storage of compliance proofs and associated metadata. The paper's primary contributions can be delineated into four components: i) presentation of a waste segregation and treatment model in Vietnam, utilizing Blockchain technology and Smart Contracts; ii) application of NFTs for storage of compliance-related content and its metadata; iii) offering a proof-of-concept implementation rooted in the Ethereum platform; and iv) executing the proposed model on four EVM and ERC721 compliant platforms, namely BNB Smart Chain, Fantom, Polygon, and Celo, to identify the most suitable platform for our proposition.

Keywords—Vietnam waste management; blockchain; smart contracts; NFT; Ethereum; Fantom; Polygon; Binance Smart Chain

I. INTRODUCTION

The challenges associated with the management and disposal of various types of waste, including domestic, industrial, and medical, pose a substantial hindrance to the economic advancement of nations [1] and a significant threat to environmental sustainability [2]. Established economies have developed stringent protocols for the inspection, categorization, and eradication of waste arising from these sources [3]. A notable portion of hazardous waste is efficiently transformed into electricity at incineration plants.

Nonetheless, in emerging economies such as the Philippines and Vietnam, where the economic potential is yet to be fully realized and the population size is considerable, methodical approaches to waste categorization and treatment have not been given the necessary attention. Predominantly, traditional waste management methods, which lack the crucial step of waste segregation at the origin sources such as residential areas, hospitals, or industrial sites, are still prevalent [4]. Investment and emphasis on pre-treatment and waste separation stages at these initial sources are considerably lacking. This neglect results in a majority of the waste being non-segregated and dumped directly into the ecosystem, leading to severe environmental pollution and contamination of surrounding areas. This waste is then conventionally collected and eliminated with no specific attention paid to the treatment of smoke and odors, thereby contributing to air and water pollution around the disposal sites.

Hazardous solid waste, for instance, rubber items like tires, and electronic components from computers and phones, need to be methodically sorted and treated via specialized procedures that safeguard the environment. To illustrate, a thermal power plant in Germany efficiently utilizes old tires as a fuel source. In order to tackle this issue, the initial step of waste segregation, also known as pre-treatment, is paramount. In the context of developed countries, industrial areas, and households, waste is typically segregated into four categories, namely i) paper-based items such as boxes and packaging; ii) recyclable waste including rubber, glass, and metal cans; iii) organic food waste; and iv) other types of waste. Each waste category is subjected to a distinctive treatment and categorization process, allowing for reuse or safe disposal to prevent harm to environmental and human health.

However, the transposition of these processes into the context of developing countries is met with challenges due to cultural differences and varying operational procedures. Notably, the constraints in waste management in these emerging economies are not merely infrastructural but also deeply rooted in public awareness and attitudes towards waste segregation and treatment. For instance, in Vietnam, wastes are often not separated and are mostly disposed of using a singular method, which involves casting them into the environment (further details on traditional waste management processes can be found in the approach section).

The conclusion of the Covid-19 pandemic has highlighted several deficiencies in global health systems [5]. Health infrastructure, medical supplies, and equipment worldwide were already strained due to a massive influx of patients, affecting the delivery of care and treatment services. This has led to the difficulty in controlling the spread of the disease, resulting in increased mortality. Particularly for emerging economies like Vietnam, which witnessed a dramatic surge in positive cases from the end of 2020 to the beginning of 2021, this issue was exacerbated by limited healthcare infrastructure, especially concerning the waste treatment process. Numerous studies, such as that conducted by [6], have concluded that unsafe procedures for handling medical waste during the pandemic have significantly contributed to the propagation of Covid-19.

To address the waste categorization and treatment problem, several models have been proposed that leverage Blockchain technology and Smart Contracts. Specific to each waste type and treatment scope, these models propose a unique treatment and transportation process for different waste types like medical waste [7], solid waste (e.g., electronic components, computers, phones) [8], household waste [9], and industrial waste [10] (see Related work for details). There is also considerable research focusing on developing waste treatment processes tailored to developing countries (e.g., India [11]; Brazil [10]; and Vietnam [12]). Regarding the Covid-19 pandemic, these approaches propose waste management and classification models for different stages (e.g., hospital/isolation zone - transportation company or transport company - waste processing company). However, these studies primarily enable governments to track and trace different types of waste (e.g., weight, waste type, etc) with ease.

In response to these issues, our goal is to design a waste categorization and treatment model rooted in Blockchain technology, Smart Contracts, and Non-Fungible Tokens (NFTs), customized to the context of Vietnam. Specifically, our proposed model empowers stakeholders to assess waste treatment levels at the output, with all relevant information processed, validated, and stored on-chain [13]. This method of on-chain storage enhances transparency compared to traditional storage methods. Additionally, we ensure system efficiency and prevent overloading through a decentralized storage approach (i.e., distributed ledger). Our model utilizes Smart Contract technology, assisting stakeholders in managing the waste treatment process, from segregation to transportation and treatment. NFT technology facilitates the storage of information regarding a garbage bag, making it easier to ascertain weight, timing, origin, and waste type (i.e., garbage, industrial, medical, domestic). Furthermore, we utilize NFTs to identify compliance or non-compliance with waste classification requirements, enabling the implementation of sanctions or rewards accordingly.

Our paper makes four primary contributions: i) we introduce a waste categorization and treatment model tailored to Vietnam, utilizing Blockchain technology and Smart Contracts; ii) we leverage NFTs to store compliance-related content and associated metadata; iii) we offer a proof-of-concept implementation based on the Ethereum platform; and iv) we execute the proposed model on four EVM and ERC721 compatible platforms, namely BNB Smart Chain, Fantom, Polygon, and Celo, to identify the most suitable platform for our proposition.

This paper comprises eight sections. After this introduction,

we delve into related works that address similar research problems. The background section considers the key technologies underpinning our work, i.e., blockchain, EVM, NFT, Smart Contracts, and the four EVM-supported blockchain platforms. We then present our approach and proposed model implementation (i.e., Execution Blueprint) in Sections IV and V. To demonstrate the efficacy of our approach, Section VI outlines our evaluation steps in various scenarios, followed by a discussion of our findings in Section VII. Finally, Section VIII provides a summary and outlines potential directions for future development.

II. RELATED WORK

Modern advancements in technology have led to the emergence of numerous innovative solutions for waste management and classification. Among these, Blockchain-based approaches have demonstrated significant potential. This section reviews a selection of notable studies in the field, broadly divided into two categories: i) Blockchain-based waste sorting and treatment solutions, and ii) region-specific waste management methodologies.

A. Blockchain-based Waste Sorting and Treatment Solutions

A plethora of studies has proposed distinctive approaches to manage different types of waste in our day-to-day lives. For instance, electronic waste (e-waste), which includes discarded electronic devices, has been addressed by Gupta et al. [8]. They proposed an Ethereum-based waste management system tailored for electrical and electronic equipment. Their model engages three key user groups: manufacturers, consumers, and retailers. Smart contracts in the system demonstrate direct constraints between these interacting entities. Retailers serve as the mediators, distributing new products to users and collecting the used items to return to manufacturers. Successful execution of these activities leads to a reward in Ether (ETH). This system eliminates the need for manufacturers to retrieve their used products directly.

In the context of solid waste, like old computers and smartphones, a model that tracks the journey of waste from its source to the treatment centers is crucial. Addressing this, Laura et al. [14] introduced a waste management system leveraging the synergy between Ethereum and QR codes. Their approach emphasizes a system that supports four stakeholders: a collection manager, a record manager, a transaction manager, and a processing manager. To determine the type of waste for disposal, each garbage bag is assigned a QR code that links to the corresponding data stored on-chain. This facilitates stakeholders to trace and ascertain the current location and estimated processing completion time of each garbage bag. By anticipating the extraction date from the garbage bag, transportation companies can determine the daily capacity for waste processing, mitigating overloading issues at waste treatment sites.

For the monitoring of cross-border waste movements in a secure, tamper-proof, and privacy-preserving manner, Schmelz et al. [15] presented an Ethereum-based study. Their approach ensures that only authorized parties can access information based on encryption technology. For authorities overseeing cross-border waste transport, they can trace location, volume,

and estimated transit times of waste units (e.g., vehicles, bags) through data stored on distributed ledgers. Shipping processes can be automated by predefined smart contracts. A limitation of this approach is the lack of mechanisms to penalize violations in waste transportation and disposal.

In another study, Francca et al. [16] proposed an Ethereum-based model for managing solid waste in small municipalities.

For models utilizing the Hyperledger Fabric platform, Trieu et al. [12] proposed a medical waste treatment model called MedicalWast-Chain. The model targets the management of medical waste from healthcare facilities, including the reuse of tools, the process of transferring medical supply waste (e.g., protective gear, gloves, masks), and waste treatment processes in factories. The objective is to enable traceability of waste origins and toxicity levels, especially crucial during a pandemic. Similarly, Ahmad et al. [17] aimed to trace personal protective equipment for healthcare workers (i.e., doctors, nurses, testers) during the pandemic. They also identified compliant and non-compliant behaviors in waste classification and collection by comparing photographs of medical waste collection sites.

To validate waste treatment processes (i.e., stakeholder interactions), Dasaklis et al. [18] proposed a blockchain-based system operable via smartphones.

B. Region-Specific Waste Management Methodologies

While Blockchain technology has proven effective in managing waste, its applications remain largely unexplored in specific regions. In light of this, we present a review of traditional waste treatment methodologies.

The efficiency of waste collection, a crucial preliminary step in waste management, is heavily influenced by the chosen travel route. Several studies have attempted to optimize this process by calculating time and cost implications (i.e., vehicle route) that influence the path of the garbage collection vehicle. Some solutions have adopted Geographic Information System (GIS) technology to manage the routes of garbage collection vehicles. For instance, Ghose et al. [19] developed a solid waste collection and treatment route for the city of Asansol in India by optimizing the path based on GIS.

On a larger scale, encompassing more than just cities, Nuortio et al. [20] proposed a well-scheduled and routed waste collection method for Eastern Finland, utilizing the neighborhood threshold metadata approach. In another example, a truck planning model for solid waste collection was proposed by Li et al. [21] for the Brazilian city of Porto Alegre.

For the European context, Gallardo et al. [22] proposed a Municipal Solid Waste (MSW) management system for the Spanish city of Castellón. The system integrated ArcGIS¹ with a planning approach to optimize travel times between locations when collecting waste in the city. This approach has shown greater efficiency compared to traditional methods [23], [24].

C. Analysis of Blockchain-based Approaches for Vietnam

The aforementioned models do not devote substantial attention to the process of recycling or refurbishing. Also, these

studies do not provide a well-rounded solution for rewarding compliance and penalizing non-compliance in waste sorting behavior of users (e.g., households, companies, businesses, or medical centers).

When considering Blockchain technology and smart contracts-based models, the existing solutions (both for Hyperledger Eco-system and Ethereum platforms) primarily focus on waste management from the initial stage up to the waste treatment plant. They lack comprehensive consideration of specific geographical characteristics, like those of Vietnam. This leaves a significant gap for a solution that incentivizes proper waste sorting habits not only for companies, businesses, medical centers but also households.

For traditional waste classification and treatment methods applied to a specific region, there has been minimal application of modern technologies to alleviate labor-intensive tasks and address current gaps (e.g., overloading, shipping process, information validation).

The present study aims to address these shortcomings. Not only do we propose a model to manage waste sorting, but we also offer a solution for rewarding compliance and handling violations of users/companies/enterprises based on NFT technology. The subsequent sections detail the background information related to our topics before elaborating on the proposed processing steps and implementation.

III. BACKGROUND AND THEORETICAL FOUNDATION

A. Blockchain Technology: An Overview

Blockchain, famously associated with Bitcoin's success [25], is a distributed ledger system. It operates on a peer-to-peer network and maintains transaction records across various computers simultaneously. This decentralized approach ensures a transparent and trustworthy data management system, eliminating the need for a central authority or intermediary for validation [26], [27], [28]. The key advantages of employing blockchain-based systems are outlined below.

- **Security:** Blockchain incorporates digital signatures and encryption to ensure a secure environment. This robust design prevents data manipulation and unauthorized access [29].
- **Fraud Prevention:** As data is replicated across multiple nodes, blockchain-based systems are resilient to hacking attempts. Moreover, the decentralized nature of blockchain allows for efficient recovery of all records [30].
- **Transparency:** With blockchain, both parties involved in a transaction receive instantaneous notifications upon completion, ensuring a seamless and reliable experience.
- **Cost-effectiveness:** Since the blockchain is a decentralized system, it bypasses intermediaries and avoids associated fees, thus reducing overall costs [31].
- **Access Control:** Blockchain provides the option to choose between a public network, accessible to all, and a permissioned network, which requires authentication for access [32].

¹A command-line based GIS system for manipulating data <https://www.arcgis.com/index.html>

- Efficiency: Transactions are processed faster in a blockchain-based system since it eliminates the need for integration with conventional payment systems [33].
- Integrity Verification: Blockchain inherently fosters a consensus-based environment, wherein the validity of participants is checked and confirmed by other network participants, further ensuring data authenticity [34].

B. Smacockchain

Smart contracts, also known as chaincode, are self-executing contracts containing the terms and conditions of an agreement directly written into code. Leveraging blockchain technology, these contracts automate transaction executions without requiring an external intervention or intermediary. Here, we delineate the salient features of smart contracts.

- Distributed: Smart contracts are replicated and distributed across all nodes of a blockchain network, distinguishing them from centralized server-based solutions.
- Deterministic: Smart contracts only perform predefined actions when specific conditions are met. Furthermore, regardless of the executor, the outcome of smart contracts remains consistent.
- Automated: Smart contracts can automate a wide range of tasks and function as self-executing programs. However, unless activated, they remain “inactive” and do not perform any action.
- Immutable: Once deployed, smart contracts cannot be altered. They can only be “deleted” if a provision for deletion was included prior to deployment, giving them an anti-forgery attribute.
- Customizable: Prior to deployment, smart contracts can be coded in different ways, making them suitable for creating diverse decentralized applications (DApps). Platforms like Ethereum are Turing complete, meaning they can solve any computational problem.
- Trust-free Environment: Smart contracts facilitate interactions between parties without requiring mutual trust, while blockchain technology ensures data accuracy.
- Transparent: Since smart contracts are based on a public blockchain, their source code remains unalterable and can be viewed by anyone.

C. Blockchain Platforms

1) *Ethereum*: Ethereum [35] is a decentralized open-source blockchain platform, renowned for its support of Turing-complete programming languages and smart contracts. It operates on the Ethereum Virtual Machine (EVM) and supports high-level programming languages such as Solidity, Serpent, LLL, and Mutan. Ethereum enables a variety of use-cases such as withdrawal limits, loans, financial contracts, and gambling markets, making it a preferred platform for smart contract development.

2) *Hyperledger fabric*: Hyperledger Fabric [36] is a permissioned, open-source, enterprise-grade distributed ledger technology (DLT) platform, tailored for large-scale commercial use. Unlike Ethereum that executes smart contracts on virtual machines, Hyperledger Fabric runs code in Docker containers, providing optimal execution speed at the expense of isolation. It supports traditional high-level programming languages such as Java and Go (Golang) over Ethereum’s exclusive smart contract languages.

D. Reasons for Choosing the Ethereum Ecosystem

The Ethereum ecosystem, underpinned by the Ethereum Virtual Machine (EVM), was chosen for our deployment due to its significant benefits. Ethereum supports smart contracts and DApps, providing a Turing-complete environment that facilitates the creation of a wide range of applications. Furthermore, Ethereum’s robust community support, rich developer tools, and high-level programming language compatibility make it a prime choice for blockchain-based development.

In addition, Ethereum’s interoperability is a significant factor. Its ecosystem includes various blockchain platforms that operate with EVM-compatible blockchains. This allows applications built on Ethereum to be easily ported to other EVM-compatible blockchains, offering flexibility in deployment options.

E. Selected Platforms for Deployment

Given the interoperability of Ethereum and the distinct advantages of EVM-compatible blockchains, we have chosen four platforms for deployment: Binance Smart Chain (BNB Smart Chain), Polygon, Fantom, and Celo.

1) *Binance smart chain*: Binance Smart Chain ² is a high-performance, low-fee blockchain platform. It supports smart contracts and is compatible with EVM, making it a viable option for deploying DApps. It also offers a dual-chain architecture with Binance Chain, allowing users to seamlessly transfer assets from one blockchain to another.

2) *Polygon*: Polygon ³ is a protocol and a framework for building and connecting Ethereum-compatible blockchain networks. It effectively transforms Ethereum into a full-fledged multi-chain system, often referred to as the “Internet of Blockchains”. Polygon combines the best of Ethereum and sovereign blockchains into a full-fledged multi-chain system.

3) *Fantom*: Fantom ⁴ is a high-performance, scalable, and secure smart-contract platform. It is designed to overcome the limitations of previous-generation blockchain platforms. Fantom’s primary proposition is its capability to perform instantaneous transactions and process large volumes at an extremely low cost, making it ideal for decentralized applications (DApps).

4) *Celo*: Celo ⁵ is an open platform that makes financial tools accessible to anyone with a mobile phone. Its mission is to build a monetary system that creates the conditions

²<https://github.com/bnb-chain/whitepaper/blob/master/WHITEPAPER.md>

³<https://polygon.technology/lightpaper-polygon.pdf>

⁴<https://whitepaper.io/document/438/fantom-whitepaper>

⁵<https://celo.org/papers/whitepaper>

of prosperity for everyone. Celo's lightweight identity and high throughput make it an optimal choice for mobile-first applications and services.

These platforms were chosen because they offer scalability, security, and efficiency while being cost-effective. Their EVM compatibility ensures smooth portability of applications built on Ethereum, providing a flexible and efficient deployment environment. For detailed understanding of these platforms, readers are referred to the respective white papers provided.

IV. APPROACH DEvised

A. Conventional Waste Management in Vietnam: An Overview

Upon surveying waste management practices across Vietnam, we find that strategies differ considerably between urban areas (cities, for example) and rural locales (Cho Lach district in Ben Tre province serves as a good case study). Urban communities, especially those densely populated, gather waste at designated spots for waste disposal firms to handle. In contrast, rural communities typically dispose of their waste directly, often impacting the natural environment adversely.

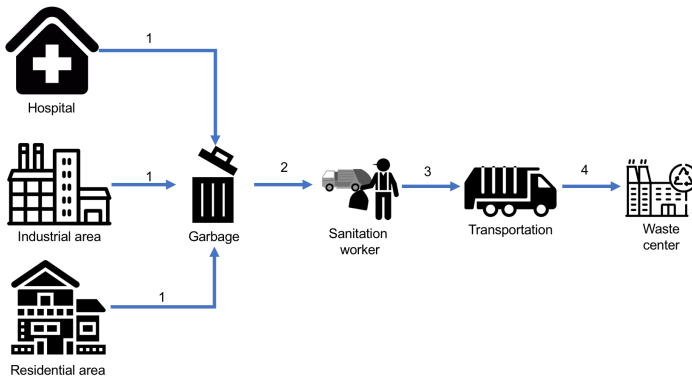


Fig. 1. Conventional waste management method in Vietnam.

Additionally, industrial and medical sectors follow their unique waste management protocols. They accumulate waste at certain locations for waste disposal companies to collect daily or semi-daily. Considering the disparities in waste segregation between urban and rural landscapes, we've formulated a waste classification and management model apt for urban settings, inspired by procedures adhered to in industrial and healthcare domains.

Fig. 1 illustrates a typical five-step waste management cycle followed in urban environments, industrial estates, and hospitals. Initially, waste is amassed at a designated location (step 1). Collection procedures (step 2) vary depending on the waste type. For instance, sanitation workers dealing with household waste (food waste) need fewer protective measures than those handling medical waste. Post collection, the waste is taken to waste sorting (step 3) and recycling centers (step 4). Depending on the waste type, these centers will either recycle or dispose of the waste (step 5).

Our study primarily focuses on the waste management process at the source (residential areas, factories, or hospitals). If individuals can segregate their waste appropriately at the source (into paper, bio, metal, and glass), it aids the subsequent

recycling and waste treatment process. However, this practice is not widely observed in Vietnam, causing difficulties for waste collectors who struggle to segregate unsorted waste. To address this, we propose using Non-Fungible Tokens (NFTs) to document instances of compliance or non-compliance with waste segregation norms. The following subsection provides a detailed description of our proposed model.

B. Waste Management Model Utilizing Blockchain Technology, Smart Contracts, and NFTs

Fig. 2 represents a six-step waste segregation and management process that integrates blockchain technology, smart contracts, and NFTs. The key distinction from the traditional model (Fig. 1) comes into play in step 2. After segregation, waste should be categorized into four groups (paper, bio, metal, and glass), each corresponding to uniquely labeled or color-coded bins.

Sanitation personnel then scrutinize the segregation process undertaken by an individual or an organization to establish whether it's compliant or non-compliant (step 3). This verification is updated onto appropriate functions within the smart contracts (step 4).

In step 5, NFTs corresponding to the individual's or organization's waste segregation actions (either compliant or non-compliant) and pertinent information (metadata; see Implementation section for additional details) are generated. Finally, the entire process is updated and archived on a distributed ledger, facilitating easy validation by concerned parties.

V. EXECUTION BLUEPRINT

Our practical model is established on two primary objectives: i) administration of data, specifically waste - originating, seeking, and revising - within a blockchain platform, and ii) fabricating Non-Fungible Tokens (NFTs) that acknowledge, reward or penalize users, which could be individuals or institutions, based on their conduct in the management and disposal of waste.

A. Origination of Data and NFTs

Fig. 3 presents the steps necessary to set up waste data. The waste data can be of different categories such as industrial, household, or medical waste. They need to be properly divided into groups like discard or repurpose, according to their toxicity levels. Detailed descriptions about each kind of waste are then affixed to each unique garbage bag.

Each bag carries a unique identifier to distinguish it by the waste type it holds. Moreover, metadata about each garbage bag is augmented to include details about the individual or organization doing the sorting, the household or company generating the waste, as well as the time and location of waste sorting. A distributed ledger-based service enables concurrent data storage from several users, hence diminishing system latency. Broadly, the waste data is structured in the following way:

```
wasteDataObject = {  
  "wasteID": industrialWasteID,  
  "sorterID": sorterID,
```

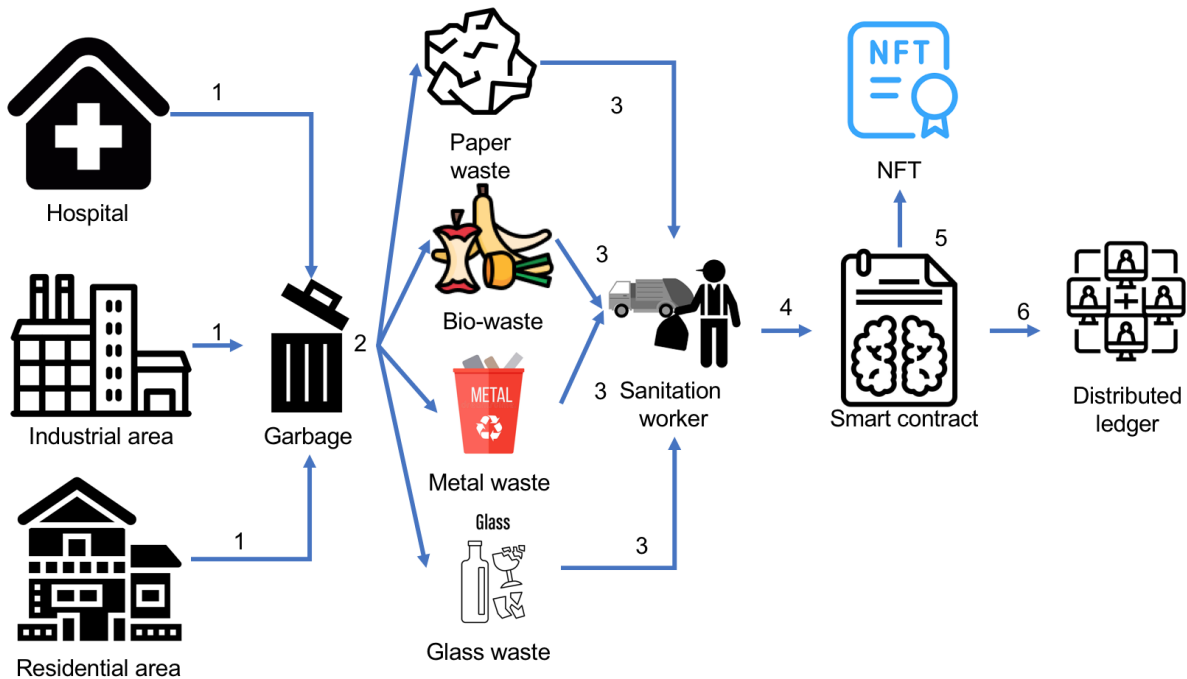


Fig. 2. Waste management framework utilizing blockchain technology, smart contracts, and NFTs.

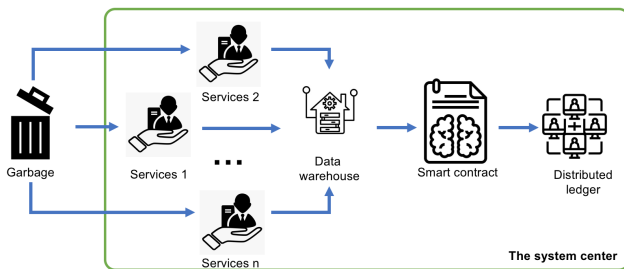


Fig. 3. Origination of data and NFTs.

```

"kind": wasteKind,
"place": place,
"amount": amount,
"unit": unit,
"bagID": bagID,
"timestamp": timestamp,
"sortingLocation": sortingLocation,
"status": null,
"repurpose": Null
};

```

Besides information about the waste’s origin, weight, and kind, the system keeps track of the status of waste bags in residential areas, factories, hospitals, etc. Specifically, the “status” switches to 1 if the relevant waste bag has been moved from its original collection location for treatment or disposal; if not, it remains at 0 (pending). The “repurpose” tag indicates 1 when the waste type is reused and 0 when pending, and is applicable to non-hazardous waste.

Post the waste sorting process, sanitation workers verify the sorted waste for compliance with set standards. The data

is then stored temporarily in a data repository, waiting for validation before being synchronized on the blockchain. This is achieved by invoking certain predefined constraints in the Smart Contract via the Application Programming Interface (API). In the process of initiating NFTs, the content of the NFT is defined as follows:

```

NFT WASTE = {
" wasteID": wasteID,
" sorterID": sorterID,
" place": place,
" bagID": bagID,
" kind": true/false,
" amount": true/false,
" timestamp": timestamp,
" inspector": cleanerID
};

```

If the values on the sorted waste bags are verified to be correct, the sorter is rewarded. In contrast, if there are discrepancies, penalties are applied. If the inspector provides incorrect information, they are the ones to be penalized.

B. Data Seeking

The data seeking procedure, much like data origination, is capable of handling multiple concurrent participants, courtesy of the distributed model on which the system operates. The services facilitate requests from the sanitation staff or any individual or organization to access the data.

The intent behind seeking data varies. Sanitation staff might be looking to review the waste sorting process or to transfer waste to the disposal companies, whereas individuals or organizations might want to gather information about the waste treatment process.

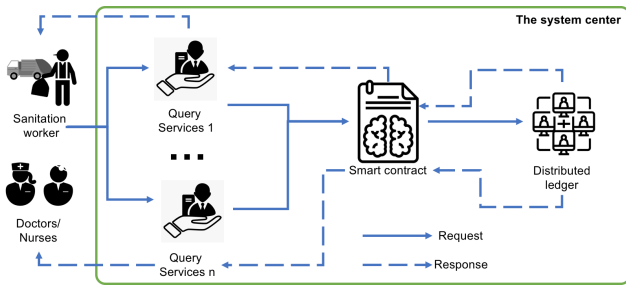


Fig. 4. Data seeking.

Fig. 4 outlines the data seeking steps. Requests are sent from the user to the smart contracts embedded within the system before retrieving data from the distributed ledger. All such requests are logged as a history for each individual or organization. If the sought information is not found (e.g., due to a wrong ID), the system responds with a 'not found' message. Regarding NFTs, all supporting services are rendered via APIs.

C. Data Revision

Data revisions are only allowed after validating that the data exists on the blockchain, following the execution of the respective data seeking procedure. In the discussion that follows, we will operate under the assumption that the sought data is indeed present on the blockchain. If not, the system will return a 'not found' message (see V-B for details).

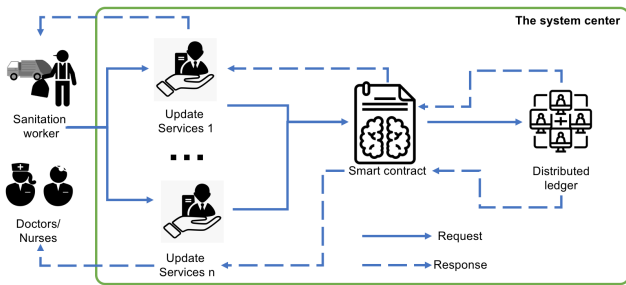


Fig. 5. Data revision.

In line with the data seeking and origination processes, the system provides revision services in the form of APIs to receive user requests before processing them via smart contracts. The primary aim of this process is to keep the time and location of waste bags updated as they move through the transportation and sorting/disposal stages. This enables the administrator to monitor the progress of waste management, right from its generation at medical centers/residential areas/factories to its final destination at waste treatment companies.

Fig. 5 outlines the waste data revision process. In the case of NFTs, the revision process only entails moving the NFT from the owner's address to a new one. If any information on an existing NFT is updated, it is stored as a new NFT (see V-A for details).

VI. DEPLOYMENT ASSESSMENT

A. Deployment Process on Four Blockchain Platforms

The deployment process of our proposed model comprises four critical steps which are applied across all four Ethereum Virtual Machine (EVM) supporting platforms (Binance Smart Chain (BNB Smart Chain), Polygon, Fantom, and Celo). These steps include:

- 1) Preliminary Setup: This step primarily involves setting up the development environment. Solidity, the programming language for Ethereum smart contracts, is used for writing the contracts. These smart contracts include rules and instructions that govern the behavior of the blockchain.
- 2) Contract Creation: Once the preliminary setup is complete, the first smart contract is created. This contract encapsulates all the rules defined in the model, such as the reward or penalty mechanism for waste sorting. The smart contract is then compiled to ensure there are no errors in the code.
- 3) NFT Generation: After the contract has been successfully compiled, it's time to generate the Non-Fungible Tokens (NFTs) that would be issued as rewards or penalties. The NFTs are created via the smart contract that has been deployed on the blockchain. Each NFT is unique and represents a real-world object, in this case, the behavior of individuals or organizations in the waste management process.
- 4) NFT Retrieval/Transfer: The final step involves updating the NFT's ownership address (i.e., transferring the NFT). This transfer is done through an operation in the smart contract. The updated NFT information is then recorded on the blockchain.

These procedures are executed in a testnet environment for each platform to evaluate their cost-effectiveness. The cost of each operation - contract creation, NFT generation, and NFT retrieval/transfer - is evaluated using the following parameters: Transaction Fee, Gas Limit, Gas Used by Transaction, and Gas Price.

B. Implementation on BNB Smart Chain (Sample Deployment)

Fig. 6 outlines the steps involved in our implementation on the BNB Smart Chain. Like the general process described above, the implementation begins with setting up the development environment and writing the contract in Solidity.

Once the smart contract is written and compiled successfully, it is deployed on the BNB Smart Chain testnet. This step creates a transaction, with details of this transaction recorded and accessible via a unique transaction hash.

Upon successful deployment of the contract, NFTs are created as per the rules defined in the smart contract. Fig. 7 shows an instance of an NFT being created.

The final step involves updating the NFT's ownership address. This involves invoking the appropriate function in the smart contract, and once executed, the NFT transfer can be seen as shown in Fig. 8.

The cost of these operations is calculated and presented in terms of the Transaction Fee, Gas Limit, Gas Used by

Txn Hash	Method	Block	Age	From	To	Value	[Txn Fee]
0xc020bd9e38391648ee...	Transfer	24862527	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	0x741c8dc8630dbde529...	0 BNB	0.00057003
0x60d184b6afb6c3fc3de...	Mint	24862522	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	0x741c8dc8630dbde529...	0 BNB	0.00109162
0x35a60f40c8a8b9d1da...	0x60806040	24862517	1 day 22 hrs ago	0xcaa9c5b45206e083f4f...	Contract Creation	0 BNB	0.02731184

Fig. 6. The transaction info on BNB smart chain.

My Name Tag:	Not Available
Contract Creator:	0xcaa9c5b45206e083f4f... at txn 0x35a60f40c8a8b9d1da...
Token Tracker:	NFT GARBAGE (GARBAGE)

Fig. 7. NFT creation on BNB smart chain.

Transaction, and Gas Price. These costs provide valuable insights into the effectiveness and efficiency of deploying our model on the BNB Smart Chain. The same deployment process and cost assessments are followed for the other platforms (Polygon, Fantom, and Celo) to evaluate their performance and cost-effectiveness. For more details, we refer the readers follow our deployment on the test-net system of the corresponding platform, namely BNB⁶; MATIC⁷; FTM⁸; and CELO⁹.

C. Transaction Fee

Table I provides a comprehensive comparison of the transaction fees incurred for various operations on the four considered blockchain platforms: BNB Smart Chain, Fantom, Polygon, and Celo.

The transaction fee is calculated for three key operations:

1) *Contract creation*: This operation involves creating and deploying the smart contract on the respective blockchain. The fee varies significantly across the platforms, with BNB Smart Chain being the most expensive at 0.02731184 BNB (approximately \$8.43). Fantom has the lowest cost for contract creation, amounting to 0.009576994 FTM (equivalent to approximately \$0.001837).

2) *Create NFT*: This operation refers to the cost of generating a Non-Fungible Token (NFT) on the blockchain. The BNB Smart Chain again appears as the most expensive

option, with a fee of 0.00109162 BNB (about \$0.34). On the contrary, the Polygon platform records the least cost, at 0.000289405001389144 MATIC (approximately \$0.00).

3) *Transfer NFT*: This refers to the cost of transferring ownership of the NFT from one address to another. BNB Smart Chain remains the most expensive platform, with a transfer fee of 0.00057003 BNB (roughly \$0.18). Conversely, the Fantom platform provides the most cost-effective solution for NFT transfer, charging a mere 0.0002380105 FTM (\$0.000046).

This table, therefore, provides a detailed overview of the cost dynamics across various platforms for different operations. BNB Smart Chain consistently shows the highest fees for all operations, while the other platforms vary in their cost-effectiveness for different operations. These insights can guide the selection of an optimal platform for deploying the recommendation model based on financial constraints and operational priorities.

D. Gas Limit

Table II presents an in-depth comparison of the gas limits on the four blockchain platforms evaluated: BNB Smart Chain, Fantom, Polygon, and Celo. The gas limit refers to the maximum amount of gas that a user is willing to spend on a transaction. Gas in blockchain is a measure of computational effort required to execute certain operations.

The table provides data for three crucial operations:

1) *Contract creation*: This column details the gas limit for creating and deploying a smart contract on the respective blockchain. Among the four platforms, Celo demands the highest gas limit for contract creation at 3,548,719, which reflects its higher computational requirements. The BNB Smart Chain has the lowest gas limit for this operation, requiring just 2,731,184.

2) *Create NFT*: This column indicates the gas limit necessary for generating a Non-Fungible Token (NFT) on each platform. Celo once again shows the highest gas limit at 142,040, demonstrating that generating an NFT on this platform is relatively computationally intensive. On the contrary, the BNB Smart Chain requires a lower gas limit, at 109,162.

3) *Transfer NFT*: This column represents the gas limit needed to transfer the ownership of an NFT from one address to another. The Celo platform necessitates the highest gas limit for NFT transfers, at 85,673, indicating a higher computational

⁶<https://testnet.bscscan.com/address/0x741c8dc8630dbde529466ecc066fe5f98b1f6ee4>

⁷<https://mumbai.polygonscan.com/address/0x3253e60880ce432dded52b5eaba9f75b92ca530a>

⁸<https://testnet.ftmscan.com/address/0x3253e60880ce432dded52b5eaba9f75b92ca530a>

⁹<https://explorer.celo.org/alfajores/address/0x3253e60880ce432DdeD52b5EAbA9f75b92Ca530A/transactions>

Txn Hash	Age	From	To	Token ID	Token
0xc020bd9e38391648ee...	1 day 22 hrs ago	0x741c8dc8630dbde529...	OUT 0xcaa9c5b45206e083f4f...	1	ERC-721: NFT....AGE
0x60d184b6afb6c3fc3de...	1 day 22 hrs ago	0x000000000000000000...	IN 0x741c8dc8630dbde529...	1	ERC-721: NFT....AGE

Fig. 8. NFT transfer on BNB smart chain.

TABLE I. TRANSACTION FEE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.02731184 BNB (\$8.43)	0.00109162 BNB (\$0.34)	0.00057003 BNB (\$0.18)
Fantom	0.009576994 FTM (\$0.001837)	0.000405167 FTM (\$0.000078)	0.0002380105 FTM (\$0.000046)
Polygon	0.006840710030099124 MATIC(\$0.01)	0.000289405001389144 MATIC(\$0.00)	0.000170007500884039 MATIC(\$0.00)
Celo	0.0070974384 CELO (\$0.004)	0.0002840812 CELO (\$0.000)	0.0001554878 CELO (\$0.000)

TABLE II. GAS LIMIT

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,184	109,162	72,003
Fantom	2,736,284	115,762	72,803
Polygon	2,736,284	115,762	72,803
Celo	3,548,719	142,040	85,673

TABLE III. GAS USED BY TRANSACTION

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	2,731,184 (100%)	109,162 (100%)	57,003 (79.17%)
Fantom	2,736,284 (100%)	115,762 (100%)	68,003 (93.41%)
Polygon	2,736,284 (100%)	115,762 (100%)	68,003 (93.41%)
Celo	2,729,784 (76.92%)	109,262 (76.92%)	59,803 (69.8%)

effort for this operation. Both BNB Smart Chain and Fantom have lower requirements, with gas limits set at 72,003 and 72,803, respectively.

This comparative data provides valuable insights into the computational demands of each blockchain platform for different operations. It highlights the variance in the computational resources needed across different platforms and operations. This information can help in selecting the most efficient platform for deploying the recommendation model based on computational and resource constraints.

E. Gas Used by Transaction

Table III provides an exhaustive analysis of the “Gas Used by Transaction” on the four blockchain platforms under examination: BNB Smart Chain, Fantom, Polygon, and Celo. This metric represents the actual amount of gas consumed to process a transaction on the blockchain.

The table breaks down the consumed gas for three different operations:

1) *Contract creation*: This is the process of deploying a smart contract on the blockchain. On BNB Smart Chain, Fantom, and Polygon, the gas used is the same as the gas limit (100%), implying that the entire computational resource allocation was utilized for this operation. However, on Celo, the gas used is 76.92% of the gas limit, suggesting a more efficient contract creation process on this platform.

2) *Create NFT*: This operation involves generating a Non-Fungible Token (NFT) on the blockchain. Again, BNB Smart Chain, Fantom, and Polygon utilize 100% of the allocated gas limit. On Celo, this operation uses 76.92% of the gas limit, indicating better computational efficiency.

3) *Transfer NFT*: This operation involves changing the ownership of an NFT from one address to another. The BNB Smart Chain platform uses 79.17% of the gas limit, while Fantom and Polygon platforms consume 93.41%. This suggests that BNB Smart Chain might be more efficient in handling NFT transfers. Conversely, Celo utilizes 69.8% of the gas limit for this operation, making it the most efficient platform among the four in terms of NFT transfer.

The table ultimately provides valuable insights into the computational efficiency of each platform. Notably, while some platforms use the entire gas limit for their operations (indicating that they are maximally utilizing the allocated resources), others use a portion of it, indicating that they are more computationally efficient. This data is critical in selecting a suitable platform for the deployment of the recommendation model, taking into account the trade-off between resource allocation and computational efficiency.

F. Gas Price

Table IV represents the “Gas Price” for executing transactions on four different Ethereum Virtual Machine (EVM) compatible blockchain platforms: BNB Smart Chain, Fantom, Polygon, and Celo. Gas prices are expressed in each blockchain platform’s native token (BNB, FTM, MATIC, and CELO, respectively) and in Gwei, where 1 Gwei equals 10^{-9} Ether.

Gas price, determined by the market conditions on the blockchain, is the cost per computational step required to execute a specific transaction or smart contract on the network.

The table breaks down the gas price for three different actions:

Contract Creation: The process of deploying a smart contract on the network. The gas prices for this action are 10 Gwei for BNB Smart Chain, 3.5 Gwei for Fantom, around 2.5 Gwei for Polygon (specifically, 2.500000011 Gwei), and 2.6 Gwei for Celo with a maximum fee per gas of 2.7 Gwei.

TABLE IV. GAS PRICE

	Contract Creation	Create NFT	Transfer NFT
BNB Smart Chain	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)	0.00000001 BNB (10 Gwei)
Fantom	0.000000035 FTM (3.5 Gwei)	0.000000035 FTM (3.5 Gwei)	0.000000035 FTM (3.5 Gwei)
Polygon	0.0000000250000011 MATIC (2.50000011 Gwei)	0.0000000250000012 MATIC (2.50000012 Gwei)	0.0000000250000013 MATIC (2.50000013 Gwei)
Celo	0.000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.000000026 CELO (Max Fee per Gas: 2.7 Gwei)	0.000000026 CELO (Max Fee per Gas: 2.7 Gwei)

Create NFT: The action of creating a Non-Fungible Token (NFT) on the network. The gas prices are identical to those for contract creation, except for Polygon, where it's slightly higher at 2.50000012 Gwei.

Transfer NFT: The operation of transferring the ownership of an NFT. Again, the gas prices are the same as for the other two operations, with the exception of Polygon, where the gas price is slightly higher at 2.50000013 Gwei.

This table is essential for understanding the costs involved in performing different actions on these platforms. It also helps in selecting a platform that balances the trade-off between computational needs and transaction costs. BNB Smart Chain has the highest gas price at 10 Gwei, while Polygon offers the most competitive gas price, hovering around 2.5 Gwei, with Celo and Fantom offering intermediate rates.

VII. DISCUSSION

A. Analysis of Transaction Costs across Different Blockchain Platforms

In our deployment assessment (VI), we detailed the transaction costs on four different EVM-enabled blockchain platforms—Binance Smart Chain (BNB), Polygon (MATIC), Fantom (FTM), and Celo (CELO)—considering three primary activities: contract creation, NFT creation, and NFT transfer. Our comprehensive examination highlighted not only the distinct monetary costs associated with each platform but also the computational costs (gas used) and gas prices.

Crucially, it is observed that the transaction value on a blockchain platform is directly influenced by the market capitalization of the platform's respective coin. As of our last observation on June 26, 2023, the total market capitalization of the four platforms—BNB, MATIC, FTM, and CELO—stood at \$50,959,673,206; \$7,652,386,190; \$486,510,485; and \$244,775,762, respectively. This market capitalization directly impacts the coin's value of each platform, although the number of coins issued at the time of system implementation is another significant factor. At the time of our evaluation, the total issuance of BNB, MATIC, FTM, and CELO was 163,276,974/163,276,974 coins; 8,868,740,690/10,000,000,000 coins; 2,541,152,731/3,175,000,000 coins; and 473,376,178/1,000,000,000 coins, respectively. Consequently, the value per coin, based traditionally on the number of coins issued and the total market capitalization, stood at \$314.98 for BNB, \$0.863099 for MATIC, \$0.1909 for FTM, and \$0.528049 for CELO.

B. Selection of Optimal Blockchain Platform for Proposed Model Deployment

Our assessments demonstrated that deploying our proposed model on Fantom offers significant advantages concerning system operating costs. Specifically, the generation and reception of NFTs incurs almost negligible fees on Fantom. Furthermore, the cost associated with creating contracts that carry a transaction execution value is extremely low, less than \$0.002.

C. Future Work

Building upon our findings, our future work will aim to implement more complex methods and algorithms, such as encryption and decryption processes, as well as more complex data structures. This will allow us to better observe the transaction costs associated with these advanced operations.

Additionally, deploying the proposed model in a real-world environment presents a compelling avenue for further research—specifically, implementing the recommendation system on the Fantom mainnet. In our current analysis, we have not taken into consideration issues related to user privacy policies, such as access control [37], [38] or dynamic policies [39], [40]. These are critical considerations that will need to be addressed in upcoming research activities.

Lastly, infrastructure-based approaches, such as gRPC [41], [42], Microservices[43], [44], dynamic transmission messages [45], and Brokerless systems [46], can be integrated into our model to enhance user interaction. For instance, we can introduce an API-call-based approach that allows for more dynamic and efficient communication between different components of the system.

VIII. CONCLUSION

In conclusion, this paper addressed the challenges of waste management and disposal in emerging economies like Vietnam by proposing a waste categorization and treatment model based on Blockchain technology, Smart Contracts, and Non-Fungible Tokens (NFTs). We highlighted the deficiencies in traditional waste management methods, particularly the lack of waste segregation and treatment at the source, leading to environmental pollution and health risks. The COVID-19 pandemic further emphasized the importance of proper waste treatment, especially in the healthcare sector. Unsafe handling of medical waste during the pandemic contributed to the spread of the disease. To address these issues, various waste management models leveraging Blockchain technology have been proposed, but they primarily focus on tracking and tracing waste rather than comprehensive waste treatment processes.

Our proposed model aims to enhance waste categorization and treatment in Vietnam by providing stakeholders with a transparent and efficient system. The use of Smart Contracts enables automated and secure waste management processes, while NFTs store essential information related to waste classification and compliance. This allows for better monitoring and implementation of sanctions or rewards based on waste management behavior.

We implemented the proposed model on four EVM-compatible platforms, namely BNB Smart Chain, Fantom, Polygon, and Celo, and evaluated their performance in terms of transaction fees, gas limits, gas used, and gas prices. Through our evaluation, we found that the Fantom blockchain platform offers the most cost-effective environment for deploying the waste management model, with negligible fees for NFT generation and low costs for contract creation. This study contributes to the field by introducing a waste categorization and treatment model customized for Vietnam and demonstrating its feasibility through a proof-of-concept implementation. The findings provide insights into the suitability of different blockchain platforms for waste management applications.

Future work includes implementing more complex methods and algorithms, considering privacy policies, and deploying the proposed model in real-world settings. Additionally, integrating infrastructure-based approaches, such as gRPC and microservices, can enhance user interaction and further optimize the waste management system.

REFERENCES

- [1] N. Singh, O. A. Ogunseitan, and Y. Tang, "Medical waste: Current challenges and future opportunities for sustainable management," *Critical Reviews in Environmental Science and Technology*, vol. 52, no. 11, pp. 2000–2022, 2022.
- [2] J. M. Chisholm, R. Zamani, A. M. Negm, N. Said, M. M. Abdel daiem, M. Dibaj, and M. Akrami, "Sustainable waste management of medical waste in african developing countries: A narrative review," *Waste Management & Research*, vol. 39, no. 9, pp. 1149–1163, 2021.
- [3] M.-G. Moldovan, D.-C. Dabija, and C. B. Pocol, "Resources management for a resilient world: A literature review of eastern european countries with focus on household behaviour and trends related to food waste," *Sustainability*, vol. 14, no. 12, p. 7123, 2022.
- [4] G. Salvia, N. Zimmermann, C. Willan, J. Hale, H. Gitau, K. Muindi, E. Gichana, and M. Davies, "The wicked problem of waste management: An attention-based analysis of stakeholder behaviours," *Journal of Cleaner Production*, vol. 326, p. 129200, 2021.
- [5] N. T. Danh, "Electronic waste classification in vietnam and some solutions to protect clean and green environment," *kalaharijournals.com*.
- [6] A. K. Das, M. N. Islam, M. M. Billah, and A. Sarker, "Covid-19 pandemic and healthcare solid waste management strategy—a mini-review," *Science of the Total Environment*, vol. 778, p. 146220, 2021.
- [7] H. Wang, L. Zheng, Q. Xue, and X. Li, "Research on medical waste supervision model and implementation method based on blockchain," *Security and Communication Networks*, vol. 2022, 2022.
- [8] N. Gupta and P. Bedi, "E-waste management using blockchain based smart contracts," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2018, pp. 915–921.
- [9] Y. Sen Gupta, S. Mukherjee, R. Dutta, and S. Bhattacharya, "A blockchain-based approach using smart contracts to develop a smart waste management system," *International Journal of Environmental Science and Technology*, vol. 19, no. 8, pp. 7833–7856, 2022.
- [10] S. Hakak, W. Z. Khan, G. A. Gilkar, N. Haider, M. Imran, and M. S. Alkathairi, "Industrial wastewater management using blockchain technology: architecture, requirements, and future directions," *IEEE Internet of Things Magazine*, vol. 3, no. 2, pp. 38–43, 2020.
- [11] A. K. Awasthi, X. Zeng, and J. Li, "Environmental pollution of electronic waste recycling in india: A critical review," *Environmental pollution*, vol. 211, pp. 259–270, 2016.
- [12] H. T. Le, K. L. Quoc, T. A. Nguyen, K. T. Dang, H. K. Vo, H. H. Luong, H. Le Van, K. H. Gia, L. V. Cao Phu, D. Nguyen Truong Quoc *et al.*, "Medical-waste chain: A medical waste collection, classification and treatment management by blockchain technology," *Computers*, vol. 11, no. 7, p. 113, 2022.
- [13] T. Hepp, M. Sharinghousen, P. Ehret, A. Schoenhals, and B. Gipp, "On-chain vs. off-chain storage for supply-and blockchain integration," *it-Information Technology*, vol. 60, no. 5-6, pp. 283–291, 2018.
- [14] M. R. Laouar, Z. T. Hamad, and S. Eom, "Towards blockchain-based urban planning: Application for waste collection management," in *Proceedings of the 9th International Conference on Information Systems and Technologies*, 2019, pp. 1–6.
- [15] D. Schmelz, K. Pinter, S. Strobl, L. Zhu, P. Niemeier, and T. Grechenig, "Technical mechanics of a trans-border waste flow tracking solution based on blockchain technology," in *2019 IEEE 35th international conference on data engineering workshops (ICDEW)*. IEEE, 2019, pp. 31–36.
- [16] A. França, J. A. Neto, R. Gonçalves, and C. Almeida, "Proposing the use of blockchain to improve the solid waste management in small municipalities," *Journal of Cleaner Production*, vol. 244, p. 118529, 2020.
- [17] R. W. Ahmad, K. Salah, R. Jayaraman, I. Yaqoob, M. Omar, and S. Ellahham, "Blockchain-based forward supply chain and waste management for covid-19 medical equipment and supplies," *Ieee Access*, vol. 9, pp. 44905–44927, 2021.
- [18] T. K. Dasaklis, F. Casino, and C. Patsakis, "A traceability and auditing framework for electronic equipment reverse logistics based on blockchain: the case of mobile phones," in *2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, 2020, pp. 1–7.
- [19] M. Ghose, A. K. Dikshit, and S. Sharma, "A gis based transportation model for solid waste disposal—a case study on asansol municipality," *Waste management*, vol. 26, no. 11, pp. 1287–1293, 2006.
- [20] T. Nuortio, J. Kytöjoki, H. Niska, and O. Bräysy, "Improved route planning and scheduling of waste collection and transport," *Expert systems with applications*, vol. 30, no. 2, pp. 223–232, 2006.
- [21] J.-Q. Li, D. Borenstein, and P. B. Mirchandani, "Truck scheduling for solid waste collection in the city of porto alegre, brazil," *Omega*, vol. 36, no. 6, pp. 1133–1149, 2008.
- [22] A. Gallardo, M. Carlos, M. Peris, and F. Colomer, "Methodology to design a municipal solid waste pre-collection system. a case study," *Waste management*, vol. 36, pp. 1–11, 2015.
- [23] F. Bonomo, G. Durán, F. Larumbe, and J. Marengo, "A method for optimizing waste collection using mathematical programming: a buenos aires case study," *Waste Management & Research*, vol. 30, no. 3, pp. 311–324, 2012.
- [24] P. Avila-Torres, R. Caballero, I. Litvinchev, F. Lopez-Irarragorri, and P. Vasant, "The urban transport planning with uncertainty in demand and travel time: a comparison of two defuzzification methods," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 3, pp. 843–856, 2018.
- [25] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," *Decentralized Business Review*, p. 21260, 2008.
- [26] H. T. Le, N. T. T. Le, N. N. Phien, and N. Duong-Trung, "Introducing multi shippers mechanism for decentralized cash on delivery system," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 6, 2019.
- [27] N. T. T. Le, Q. N. Nguyen, N. N. Phien, N. Duong-Trung, T. T. Huynh, T. P. Nguyen, and H. X. Son, "Assuring non-fraudulent transactions in cash on delivery by introducing double smart contracts," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 5, pp. 677–684, 2019.
- [28] X. S. Ha, T. H. Le, T. T. Phan, H. H. D. Nguyen, H. K. Vo, and N. Duong-Trung, "Scrutinizing trust and transparency in cash on delivery systems," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2020, pp. 214–227.

- [29] N. Duong-Trung, X. S. Ha, T. T. Phan, P. N. Trieu, Q. N. Nguyen, D. Pham, T. T. Huynh, and H. T. Le, "Multi-sessions mechanism for decentralized cash on delivery system," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, 2019.
- [30] N. Duong-Trung, H. X. Son, H. T. Le, and T. T. Phan, "Smart care: Integrating blockchain technology into the design of patient-centered healthcare systems," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*, ser. ICCSP 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 105–109.
- [31] —, "On components of a patient-centered healthcare system using smart contract," in *Proceedings of the 2020 4th International Conference on Cryptography, Security and Privacy*. New York, NY, USA: Association for Computing Machinery, 2020, p. 31–35.
- [32] H. T. Le, T. T. L. Nguyen, T. A. Nguyen, X. S. Ha, and N. Duong-Trung, "Bloodchain: A blood donation network managed by blockchain technologies," *Network*, vol. 2, no. 1, pp. 21–35, 2022.
- [33] N. T. T. Quynh, H. X. Son, T. H. Le, H. N. D. Huy, K. H. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, N. Duong-Trung *et al.*, "Toward a design of blood donation management by blockchain technologies," in *International Conference on Computational Science and Its Applications*. Springer, 2021, pp. 78–90.
- [34] H. T. Le, L. N. T. Thanh, H. K. Vo, H. H. Luong, K. N. H. Tuan, T. D. Anh, K. H. N. Vuong, H. X. Son *et al.*, "Patient-chain: Patient-centered healthcare system a blockchain-based technology in dealing with emergencies," in *International Conference on Parallel and Distributed Computing: Applications and Technologies*. Springer, 2022, pp. 576–583.
- [35] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, 2020.
- [36] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, Y. Manevich *et al.*, "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the thirteenth EuroSys conference*, 2018, pp. 1–15.
- [37] H. X. Son, M. H. Nguyen, H. K. Vo *et al.*, "Toward an privacy protection based on access control model in hybrid cloud for healthcare systems," in *International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on European Transnational Education (ICEUTE 2019)*. Springer, 2019, pp. 77–86.
- [38] H. X. Son and N. M. Hoang, "A novel attribute-based access control system for fine-grained privacy protection," in *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, 2019, pp. 76–80.
- [39] S. H. Xuan, L. K. Tran, T. K. Dang, and Y. N. Pham, "Rew-xac: an approach to rewriting request for elastic abac enforcement with dynamic policies," in *2016 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 2016, pp. 25–31.
- [40] H. X. Son, T. K. Dang, and F. Massacci, "Rew-smt: a new approach for rewriting xacml request with dynamic big data security policies," in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, 2017, pp. 501–515.
- [41] L. T. T. Nguyen *et al.*, "Bmdd: a novel approach for iot platform (broker-less and microservice architecture, decentralized identity, and dynamic transmission messages)," *PeerJ Computer Science*, vol. 8, p. e950, 2022.
- [42] L. N. T. Thanh *et al.*, "Toward a security iot platform with high rate transmission and low energy consumption," in *International Conference on Computational Science and its Applications*. Springer, 2021.
- [43] —, "Toward a unique iot network via single sign-on protocol and message queue," in *International Conference on Computer Information Systems and Industrial Management*. Springer, 2021.
- [44] L. N. T. Thanh, N. N. Phien, T. A. Nguyen, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, and H. X. Son, "Ioht-mba: An internet of healthcare things (ioht) platform based on microservice and brokerless architecture," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 7, 2021. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2021.0120768>
- [45] L. N. T. Thanh *et al.*, "Uip2sop: A unique iot network applying single sign-on and message queue protocol," *IJACSA*, vol. 12, no. 6, 2021.
- [46] L. N. T. Thanh, N. N. Phien, H. K. Vo, H. H. Luong, T. D. Anh, K. N. H. Tuan, H. X. Son *et al.*, "Sip-mba: A secure iot platform with brokerless and micro-service architecture," 2021.

Deep Learning-based Sentence Embeddings using BERT for Textual Entailment

Mohammed Alsuhaibani

Department of Computer Science, College of Computer,
Qassim University, Buraydah 52571, Saudi Arabia

Abstract—This study directly and thoroughly investigates the practicalities of utilizing sentence embeddings, derived from the foundations of deep learning, for textual entailment recognition, with a specific emphasis on the robust BERT model. As a cornerstone of our research, we incorporated the Stanford Natural Language Inference (SNLI) dataset. Our study emphasizes a meticulous analysis of BERT’s variable layers to ascertain the optimal layer for generating sentence embeddings that can effectively identify entailment. Our approach deviates from traditional methodologies, as we base our evaluation of entailment on the direct and simple comparison of sentence norms, subsequently highlighting the geometrical attributes of the embeddings. Experimental results revealed that the L_2 norm of sentence embeddings, drawn specifically from BERT’s 7th layer, emerged superior in entailment detection compared to other setups.

Keywords—Textual entailment; deep learning; entailment detection; BERT; text processing; natural language processing systems

I. INTRODUCTION

Textual entailment (TE), an essential notion within natural language processing (NLP), is expressed as a binary correlation between two segments of text [1]. Text T is stated to entail another text H if the comprehension gathered from T would compel a reader to deduce that H is most probable [2]. For example, the sentence “The dog is playing in the park” entails that “There is a dog at the park”. This unfolds as a unidirectional correlation, where TE serves as a fundamental pillar within NLP, supporting numerous applications in various disciplines.

TE’s multifaceted applications extend across diverse tasks, including question answering (QA) [3], where the precise extraction of responses from intricate texts hinges significantly on the correct discernment of entailment. It also impacts the effectiveness of information retrieval (IR) [4] tasks and the success of information extraction processes. TE is also an essential ingredient in the creation of text summarization [5, 6] mechanisms. The vast reach of these applications accentuates the crucial nature of textual entailment and the importance of its accurate identification.

Nonetheless, TE introduces a notable challenge, especially in terms of understanding the semantic relationships between sentences [7, 8, 9]. To tackle this, sentence embeddings have garnered significant attention lately. At their core, sentence embeddings are condensed vector depictions of sentences created to encode their semantic meanings within a fixed-dimensional vector [10]. The deployment of sentence embeddings enables swift and effective comparison and assessment of different

sentences, acting as an important instrument in a range of NLP tasks, including TE.

In the domain of sentence embeddings generation, deep learning has led the advancements. The hierarchical learning aptitudes of deep learning models enable them to produce semantically rich sentence embeddings, encompassing the intricate syntactic and semantic attributes of sentences. Notably, these models have demonstrated remarkable proficiency in discerning nuanced relationships, like entailment, among sentences [11, 12].

In recent advancements of deep learning for NLP, Transformer-based models, with particular emphasis on BERT (Bidirectional Encoder Representations from Transformers) [13], have signified noteworthy progress. The ability of BERT to consider the complete context of a sentence bi-directionally (left and right) permits the creation of superior-quality sentence embeddings. This unique capability has earned BERT widespread recognition and usage in the NLP community, particularly for tasks such as TE [14, 15, 16].

The assessment of various methods and models in TE rests on numerous specific datasets. The Stanford Natural Language Inference (SNLI) [1] dataset is one such resource, offering a large collection of sentence pairs annotated for entailment, contradiction, and neutrality. Resources like SNLI enable consistent and comparable evaluation of different TE techniques, encouraging advancement in the field.

Despite the remarkable progress in TE, current methods, especially those founded on deep learning, still exhibit shortcomings. These include an intense dependence on complex architectural designs and extensive computational resources. In addition, a majority of these models primarily concentrate on the syntactic features of sentences, frequently neglecting the geometric attributes of sentence embeddings.

To address these issues, our study delves into the detailed examination of the use of sentence embeddings for TE. Utilizing, directly, the strength of the BERT model, we scrutinize the effects of employing varying layers for the extraction of sentence embeddings. Our study departs from traditional methods by assessing entailment through the comparison of sentence norms, thereby focusing on the geometric characteristics of the embeddings, a less explored yet potentially beneficial aspect.

Our hands-on findings underline the good performance of the L_2 norm of sentence embeddings, specifically those extracted from the 7th layer of BERT. These findings offer a fresh perspective on the TE. Our results particularly emphasise the importance of layer selection in the extraction of sentence

embeddings as well as the consideration of the geometric properties of sentence embeddings in addressing TE.

The remainder of this paper unfolds as follows. We will first dive into the related work in Section II, where we discuss the key literature on textual entailment and sentence embeddings. In Section III we will share our proposed method which utilizes the BERT model. Next, in Section IV, we will discuss SNLI dataset that we used for our experiments. We then move to the experiments and results in Section V, where we lay out the outcomes of the experiments and interpret our results and speak on any limitations we have come across. And lastly, in the conclusion, Section VI, we will bring everything together by summarizing our findings, reaffirming what our study brings to the field, and pondering over potential areas for future research.

II. BACKGROUND AND RELATED WORK

Textual Entailment (TE), also known as Natural Language Inference (NLI), entails determining the relationship between two sentences, specifically, if one sentence (the hypothesis) implies, contradicts, or remains neutral to the other (the premise) [17]. This is a demanding task as it necessitates understanding the essence of both sentences and their interplay.

One method to accomplish TE employs sentence embeddings, which are vector representations encapsulating the semantic significance of sentences [18]. These embeddings can be used to train a model to anticipate the relationship dynamics between a pair of sentences.

There exists a plethora of techniques to generate sentence embeddings. A prevalent approach involves deploying a word embedding model to create word embeddings [19, 20, 11], which are then amalgamated to craft a sentence embedding. An alternative strategy employs a deep learning model specifically trained for generating sentence embeddings [21, 22].

BERT [13] has gained popularity as a deep learning model for sentence embeddings. As a transformer-based model, BERT is pre-trained on an extensive corpus of text, enabling it to effectively learn and represent word and sentence meanings. This capability is useful for a wide spectrum of NLP tasks, including TE. There has been a growing body of research on using BERT for TE. In fact, when Devlin et al. introduced BERT itself, it was trained using next-word prediction and missing-word prediction, allowing it to acquire meaningful word and sentence representations and has proven useful for several NLP tasks, including TE.

Moreover, Lin and Su [15] examine BERT's proficiency in handling TE tasks, particularly its capability to bypass any latent biases in the dataset. To simplify the investigation, they design a straightforward entailment judgment scenario using only binary predicates in clear English. The results suggest that BERT's learning curve is somewhat slower than expected. However, they found that incorporating task-specific features significantly improved the learning efficiency, leading to a data reduction by a factor of 1,500. This key discovery highlights the importance of domain knowledge in effectively utilizing neural networks for TE tasks.

Similarly, Gajbhiye et al. [23] introduce a new model for TE, dubbed External Knowledge Enhanced BERT (ExBERT).

It improves BERT's language understanding and reasoning capabilities by integrating commonsense knowledge from external sources into the existing contextual representation. The model uses BERT-derived contextual word representations to pull and encode relevant knowledge from knowledge graphs. It's designed to seamlessly blend this external knowledge into the reasoning process.

Pang et al. [24] have developed a method for integrating syntax into TE models. Their approach uses contextual token-level vector representations derived from a pre-trained dependency parser. This technique, similar to other contextual embedders, can be applied to a wide range of neural models. They tested this method with some established TE models, such as BERT. The findings showed an increase in accuracy across the benchmark datasets.

Cabezudo et al. [25] investigate various methods to enhance inference recognition in the ASSIN [26] dataset, a dataset specifically designed for entailment recognition in Portuguese. They also study the effects of adding external data, such as multilingual data or an automatically translated corpus, to improve model training. They use the multilingual pre-trained BERT model in their experiment and their results show an improvement in the ASSIN. Interestingly, their findings suggest that using external data does not significantly improve the performance of the model.

Wehnert et al. [27] have introduced three distinct methods for the classification of entailment. The first approach harmonizes Sentence-BERT embeddings with a graph neural network, while the second strategy leans on the specific LEGAL-BERT model, which undergoes additional training on the competition's retrieval task and is fine-tuned specifically for entailment classification. Their third method ingeniously employs the KERMIT encoder to embed syntactic parse trees and integrates this with a BERT model. Their study delves into the potential of this third tactic and provides insights into why the LEGAL-BERT submissions, among all entries, might have managed to edge out the graph-based method in performance.

Shajalal et al. [28] develop a new method for identifying the textual entailment relationship between a text and its hypothesis. They introduce a new semantic feature that uses empirical threshold-based semantic text representation. This approach makes use of an element-wise Manhattan distance vector-based feature, designed to understand the semantic entailment relationship within a text-hypothesis pair. They tested their method using several experiments on the benchmark entailment classification dataset, SICK-RTE [29], with a variety of machine learning algorithms. Their empirical sentence representation technique improved the semantic understanding of the texts and hypotheses.

Jiang and de Marneffe [30] have taken on the task of addressing an issue prevalent in TE datasets. They have come up with a strategy, redefining the use of the CommitmentBank for TE. Their idea is to adjust the emphasis on how committed a speaker is to the complements of clause-embedding verbs in a range of contexts that cancel entailment. This move leads to the creation of hypotheses that are free from artefacts and naturally intertwined with the premises. Even though their fresh approach lets a BERT-based model hit a good result with BERT, they stated that the model is not yet fully grasping

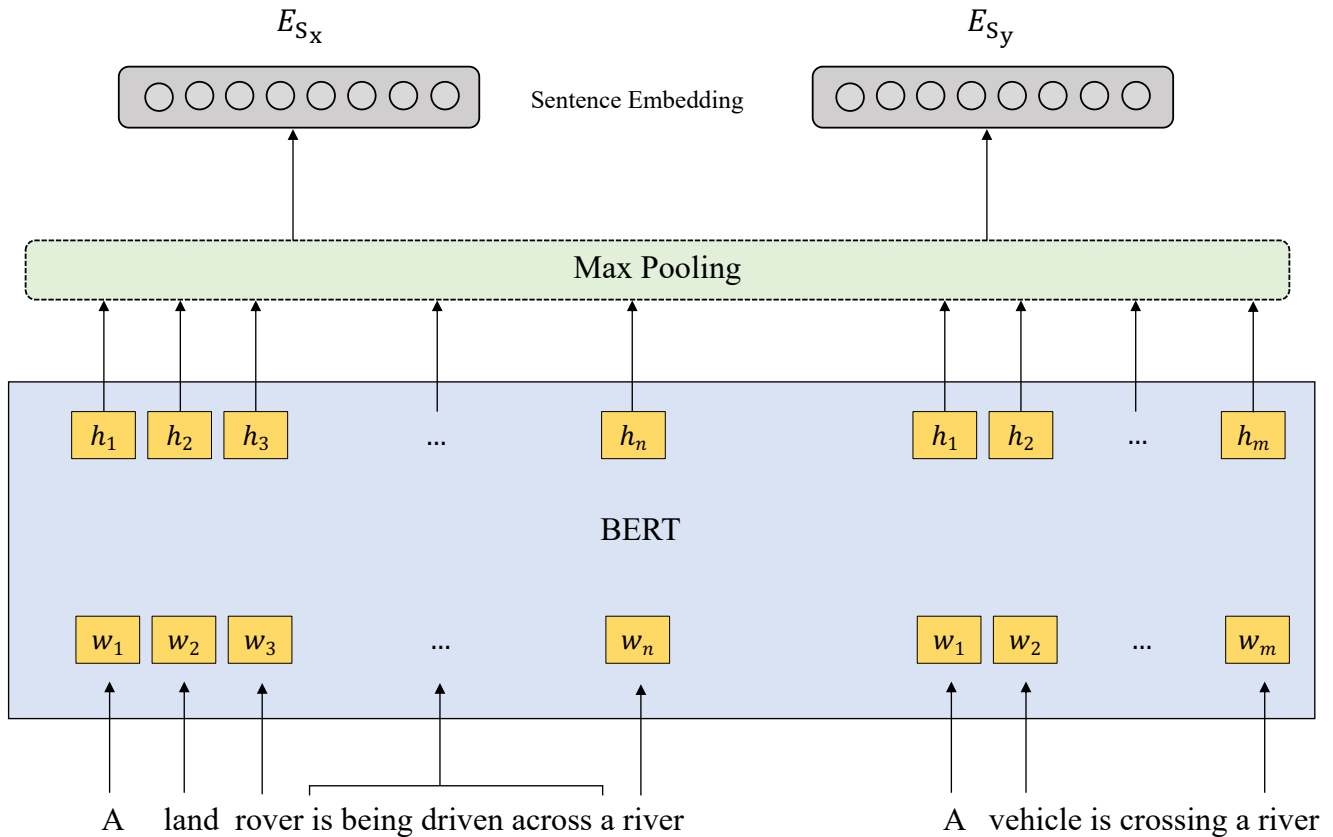


Fig. 1. Extraction of sentence embeddings from BERT with a *max* pooling strategy from the token-level embeddings [13].

the nuances of pragmatic reasoning and certain linguistic generalizations.

While the above-mentioned approaches significantly advanced TE, its reliance on intricate designs and significant computational power is notable. To rectify this, our research investigates the application of sentence embeddings in TE. We simply and directly utilize the BERT model’s potential, exploring the effects of various layers for sentence embeddings extraction. In contrast to conventional approaches, we utilize a simple and straightforward approach to evaluate entailment by comparing sentence norms, spotlighting the geometric aspects of embeddings, a relatively uncharted but potentially advantageous area.

III. PROPOSED APPROACH

Our approach to TE revolves around using BERT to extract sentence embeddings. While loading pre-trained BERT and tokenizer, we set configurations for sub-token pooling, which determines the token piece embeddings used in constructing the token embedding. Options include using the *first* subtoken, the *last* subtoken, both the *first* and *last*, or an *average* overall (mean). Additionally, we specify the layer (layers 1 to 12) from which the embeddings should be extracted. Specifically, as shown in Fig. 1, we generate sentence embeddings for each pair of sentences in the dataset. We will feed the premise and hypothesis into BERT and extract the output of the [CLS] special token, which is a fixed-length representation of the

entire input sequence. This will provide us with a pair of sentence embeddings that capture the semantic and syntactic information of the premise and hypothesis.

Given a pair of sentences (x, y) with $x = w_1, \dots, w_n$ and $y = w_1, \dots, w_m$ forming a tuple, we use the loaded pre-trained BERT model to encode each sentence individually. We employed two possible strategies: default document embeddings and token-based document embeddings.

In default document embeddings, we derive one vector representing the entire sentence as $E_{S_i} = TN(i)$, where $i \in x, y$ and TN denotes a Transformer-based network, BERT. Basically, it extracts one feature as the sentence embedding using a default pooling strategy that simply selects the first token feature [CLS] from the standard word-piece tokenization as proposed in BERT. On the other side, in the token-based document embeddings (Fig. 1), we extract a vector corresponding to each token in a sentence, for example, $S_x = (E_1, \dots, E_n)$, where $E_i = TN(w_i) \in \mathbb{R}^D$ (D is the embedding size). To generate a sentence vector, we then compute either a *min*, *max* or *mean* pool across all these token vectors.

$$E_{S_x} = \frac{1}{n} \sum_i^n E_i \quad (1)$$

When using the *mean*, we calculate an average across all vectors to derive a sentence vector. The sentence embedding of x , E_{S_x} , is calculated using (1), and E_{S_y} for y is computed

similarly. Besides the *mean*, we also test with other pooling strategies like *min* and *max*. *Min* involves sorting the token vectors based on their norm magnitude and using the vector with the least magnitude as the sentence vector. Conversely, *max* employs the vector with the largest norm magnitude as the sentence vector. We will use the *max* pooling in our experiments which empirically gives the best performance as we will detail in Section V.

Lastly, we predict entailment by comparing the norms of the pair of sentences in our input tuple. If the norm of x is greater than or equal to the norm of y , we consider it as entailment; otherwise, it is not (as shown in (2) and (3)). This approach provides a direct and effective way to determine TE.

$$V = \|E_{S_x}\|_2 \geq \|E_{S_y}\|_2 \quad (2)$$

$$\text{Entailment} = \begin{cases} \text{True} & \text{if } V(x, y), \\ \text{False} & \text{otherwise.} \end{cases} \quad (3)$$

IV. DATA

As a main dataset, we have leveraged the Stanford Natural Language Inference (SNLI) [1] dataset, a comprehensive collection of sentence pairs instrumental in training TE models. The SNLI is a robust dataset of approximately 570,000 human-authored English sentence pairs, each meticulously annotated to ensure balanced classification across three categories: entailment, contradiction, and neutral. Its wide acceptance and usage for training and testing models in TE have earned it the reputation of a standard benchmark within the field. It is worth noting that the creation of this dataset involved a crowdsourcing approach. Meaning human contributors generated the sentence pairs and assigned the entailment categories. This human involvement ensures the quality and reliability of the data.

SNLI dataset has been a pivotal element in the evolution of many contemporary NLP models, including transformative models like BERT and their subsequent iterations.

Table I features select examples from the SNLI dataset used in our approach. For ease of comprehension, we've adopted a color-coding scheme: instances of entailment are presented in green-shaded rows, neutral examples have been uncolored, while contradiction cases appear in rows shaded red. This approach to color differentiation offers an intuitive visualization of the varied sentence pairs that the SNLI dataset encompasses.

V. EXPERIMENTS AND RESULTS

A. Experimental Settings

In this section, we provide an outline of the steps we have followed to execute our experiments, covering the specific details of loading data, data preprocessing, and the application of pre-trained models and tokenizers.

Our experimental framework incorporates the use of the Hugging Face API¹ for the purpose of loading BERT pre-trained model and tokenizers. As part of our configuration

parameters, we have included a setting for sub-token pooling. This setting dictates the manner in which token piece embeddings are utilized to form the final token embedding.

The data loading process involves drawing sentences from one of two file formats: Excel (*.xlsx*) or JavaScript Object Notation (*.json*). Furthermore, we have prepared an alternate method to load data, using the Hugging Face dataset loader object as a substitute for traditional content loading from text or *json* files.

In the data preprocessing step, we apply a series of operations to refine and structure the data. Initially, we clean each sentence pair in the dataset by eliminating superfluous spaces found at the sentence boundaries. Following this, we organize the cleaned pairs of sentences into tuples, i.e., a sentence pair (*sentence1, sentence2*), culminating in a list of such tuples. This process ensures that our data is well-organized and conducive to subsequent tasks.

With the aid of the Hugging Face API, we have streamlined the process of loading BERT pre-trained weights for a variety of PyTorch² and TensorFlow³ models. This step is critical in harnessing the capabilities of BERT pre-trained model, which has already acquired useful representations from extensive text corpora, to kickstart our task-specific model.

Subsequent to extracting a vector that corresponds to each token in a sentence, we carry out additional processing on these token vectors to derive a unified sentence vector. As highlighted in Section III, this is achieved by implementing one of the multiple pooling strategies, *min*, *max* or *mean* across all token vectors.

Our initial experimentation revealed that the *max* pooling strategy surpassed the performance offered by the *min* and *mean* strategies. Hence, we chose to incorporate the *max* pooling strategy in all subsequent experiments for generating sentence vectors from token vectors. This choice proved pivotal in boosting the effectiveness of our entailment detection procedure.

Alongside our selected pooling strategy, we also examined the effect of different layers within the BERT model on our results. We extracted embeddings from a range of layers within BERT, extending from layer 1 to layer 12, and studied their influence on the task of TE. This experiment offers insight into the role each layer has in shaping the quality of sentence embeddings. This expansive exploration across all layers of the BERT model enables us to pinpoint the optimal layer for our specific task, a factor in boosting the efficacy of our entailment detection procedure.

In an extension to our experimental setup, we investigated the impact of various norms, L_1 , L_2 , and L_{∞} on the entailment detection. As norms play a vital role in comparing sentence embeddings in our methodology, experimenting with different norms helped us identify which norm leads to the most precise and reliable entailment predictions. The outcomes of these investigations are reported in our study, shedding light on the influence of each norm on the performance of our entailment detection approach.

¹<https://huggingface.co/models>

²<https://pytorch.org/>

³<https://www.tensorflow.org/>

TABLE I. RANDOMLY CHOSEN SAMPLES FROM THE SNLI DATASET USED IN THE PROPOSED APPROACH, COLOR-CODED BY ENTAILMENT CATEGORY

Text	Judgments	Hypothesis
A middle-aged man in a gray t-shirt and brown pants sitting on his bed reading a flyer-like paper.	entailment E E E E E	A man is sitting on his bed reading.
A young boy and girl playing baseball in a grassy field.	entailment N E E E E	Kids play baseball.
Numerous people sitting in a dim lit room talking, drinking coffee and using computers.	entailment E E E E E	People are in a dimly lit room drinking coffee.
A white race dog wearing the number eight runs on the track.	entailment E E E E E	A dog is running.
A woman reaching for candy bars that are on a shelf.	neutral N N E N C	The candy bars are above the womans head.
The boy wearing the blue hooded top is holding a baby goat in his arms.	neutral N C N N N	The goat jumped into the boys arms.
A little girl is sitting on a bench in a park.	neutral N N N N N	The little girl is having fun.
A small child playing in a dusty square.	neutral E N N N N	A child is playing with a doll.
Multiple people starting to pack their parachutes after a successful skydive.	contradiction C C C C C	cat chased by tiger.
A swimming dog with a small branch in its mouth.	contradiction C C C C N	A dog is ice skating.
A man with a mustache is playing ice hockey with snow in the background.	contradiction C C C C C	People are swimming in the lake.
A busy street full of shops and people holding hands and walking.	contradiction C C C C C	People sitting in a restaurant.

In Section III, we laid out our strategy for evaluating the proposed method, which, despite its apparent simplicity, yields potent results. The heart of our approach to entailment prediction lies in comparing the norms of the sentence pairs that make up our input tuple. If the norm of x equals or surpasses that of y , we mark it as an entailment instance. In contrast, if it fails to meet this criterion, we label it as non-entailment (refer to Equations (2) and (3) for further clarity). When it comes to gauging performance, we turn to the accuracy metric. This indicator gives us the ratio of successful classifications. By resorting to this measure, we can quantify how adept our model is at correctly categorizing sentence pairs in alignment with their actual entailment status. This simple yet effective measure offers a clear insight into our proposed approach's efficiency in entailment prediction.

B. Results and Discussion

The results reflected in Table II offer a thorough perspective of the outcomes generated through our proposed approach. We have incorporated accuracy percentages that depict the reper-

cussions of diversifying two primary parameters: the BERT model's layers (from 1 to 12) and the types of norms ($L1$, $L2$, and L -inf). Regardless of these alterations, the max pooling strategy remained a constant, thereby offering a consistent benchmark for comparison.

Our findings lead us to two insights. The first is related to the choice of norm type; the $L2$ norm systematically outpaced both $L1$ and L -inf norms regardless of the layer, and $L1$ come second. Whereas, L -inf performs poorly across the layers.

Our second insight arises from the analysis of BERT model's layers. As per the empirical findings, it appears that the 7th layer offers an optimal environment for the extraction of embeddings with as high accuracy as %91. This is important as it aids us in pinpointing the most suitable layer, thereby optimizing the sentence embedding generation process.

To simplify the understanding of the results and make them visually discernible, we have plotted the model's performance. For this, in Fig. 2, we considered the $L2$ norm (proven to offer superior results) and plotted its influence on the

TABLE II. PERFORMANCE ACCURACY OF THE PROPOSED APPROACH WITH BERT LAYER VARIATION AND NORM TYPES WITH MAX POOLING STRATEGY. BOLD INDICATES THE BEST PERFORMANCE FOR EACH NORM

Norm	Layers											
	1	2	3	4	5	6	7	8	9	10	11	12
L2	0.75	0.83	0.83	0.84	0.82	0.83	0.91	0.87	0.77	0.77	0.76	0.83
L1	0.73	0.81	0.81	0.81	0.77	0.74	0.80	0.83	0.65	0.60	0.57	0.59
L-inf	0.26	0.22	0.17	0.16	0.19	0.24	0.33	0.22	0.47	0.41	0.32	0.49

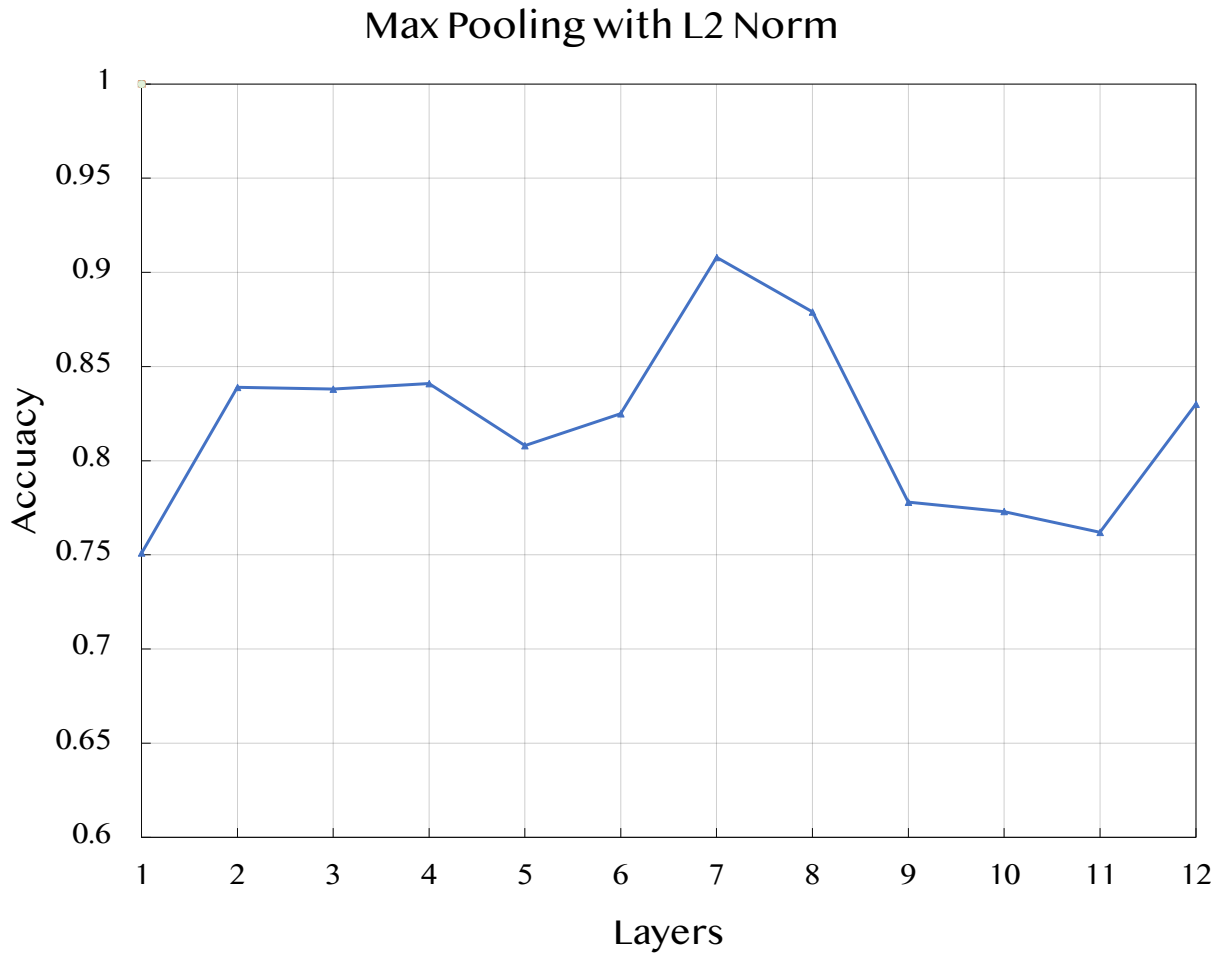


Fig. 2. Proposed approach performance with L2 norm and various layers.

varying layers, with the latter serving as the x -axis. Despite the changes in layers, we ensured the max pooling strategy remained unchanged, facilitating a focused study on the layers' influence. The resulting graph offers a straightforward visual comparison of the performance impact due to different layers.

VI. CONCLUSION

In this study, we have delved TE, using the expansive SNLI dataset as our sandbox. Our approach lies in leveraging the strength of pre-existing models, with an emphasis on the BERT model. Our methodology consists of extracting token embeddings and transforming them into sentence vectors. In our quest to streamline these vectors, we experimented with

several pooling strategies, min , max , and $mean$. Our observations consistently pointed towards the max pooling strategy as the most effective. We focused on the implications of various layers within the BERT model on the task of entailment detection. Our experiments revealed that the seventh layer of the model stood out as the most impactful for generating potent embeddings for this task.

Norms, too, were given considerable attention in our experimental setup. We tested different norms, namely $L1$, $L2$, and L -inf. Our findings tipped the scales in favor of the $L2$ norm, emphasizing the influential role norms play in determining the quality of entailment detection.

To sum it up, our research presents a direct and simple approach for effective entailment detection by utilizing BERT. It underscores the importance of which layers to select for extracting embeddings, the pooling strategies to implement, and the norms to use. Future exploration could include testing our approach on other pre-trained models and entailment datasets to enhance its generalizability.

REFERENCES

- [1] S. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 632–642.
- [2] Q. Chen, X. Zhu, Z.-H. Ling, S. Wei, H. Jiang, and D. Inkpen, "Enhanced lstm for natural language inference," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2017.
- [3] X. Wang, P. Kapanipathi, R. Musa, M. Yu, K. Talamadupula, I. Abdelaziz, M. Chang, A. Fokoue, B. Makni, N. Mattei *et al.*, "Improving natural language inference using external knowledge in the science questions domain," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 7208–7215.
- [4] K. Zhou, Q. Qiao, Y. Li, and Q. Li, "Improving distantly supervised relation extraction by natural language inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 14 047–14 055.
- [5] H. Song, W.-N. Zhang, J. Hu, and T. Liu, "Generating persona consistent dialogues by exploiting natural language inference," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 05, 2020, pp. 8878–8885.
- [6] H. Chouikhi, M. Alsuhaibani, and F. Jarray, "Bert-based joint model for aspect term extraction and aspect polarity detection in arabic text," *Electronics*, vol. 12, no. 3, p. 515, 2023.
- [7] H. Choi, J. Kim, S. Joe, and Y. Gwon, "Evaluation of bert and albert sentence embedding performance on downstream nlp tasks," in *2020 25th International conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 5482–5487.
- [8] Z. Chen, Q. Gao, and L. S. Moss, "Neurallog: Natural language inference with joint neural and logical reasoning," in *Proceedings of* SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics*, 2021, pp. 78–88.
- [9] A. Talman, A. Yli-Jyrä, and J. Tiedemann, "Sentence embeddings in nli with iterative refinement encoders," *Natural Language Engineering*, vol. 25, no. 4, pp. 467–482, 2019.
- [10] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3982–3992.
- [11] J. Maillard, S. Clark, and D. Yogatama, "Jointly learning sentence embeddings and syntax with unsupervised tree-lstms," *Natural Language Engineering*, vol. 25, no. 4, pp. 433–449, 2019.
- [12] T. Gao, X. Yao, and D. Chen, "Simcse: Simple contrastive learning of sentence embeddings," in *2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*. Association for Computational Linguistics (ACL), 2021, pp. 6894–6910.
- [13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [14] Y. Song, J. Wang, Z. Liang, Z. Liu, and T. Jiang, "Utilizing bert intermediate layers for aspect based sentiment analysis and natural language inference," *arXiv preprint arXiv:2002.04815*, 2020.
- [15] Y.-C. Lin and K.-Y. Su, "How fast can bert learn simple natural language inference?" in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 626–633.
- [16] Q. He, H. Wang, and Y. Zhang, "Enhancing generalization in natural language inference by syntax," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 4973–4978.
- [17] I. Androutsopoulos and P. Malakasiotis, "A survey of paraphrasing and textual entailment methods," *Journal of Artificial Intelligence Research*, vol. 38, pp. 135–187, 2010.
- [18] J. Yu and J. Jiang, "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 236–246.
- [19] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [20] M. Alsuhaibani, D. Bollegala, T. Maehara, and K.-i. Kawarabayashi, "Jointly learning word embeddings using a corpus and a knowledge base," *PLoS one*, vol. 13, no. 3, p. e0193094, 2018.
- [21] M. Pagliardini, P. Gupta, and M. Jaggi, "Unsupervised learning of sentence embeddings using compositional n-gram features," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, 2018.
- [22] S. Arora, Y. Liang, and T. Ma, "A simple but tough-to-beat baseline for sentence embeddings," in *5th International Conference on Learning Representations, ICLR 2017*, 2019.
- [23] A. Gajbhiye, N. A. Moubayed, and S. Bradley, "Exbert: An external knowledge enhanced bert for natural language inference," in *Artificial Neural Networks and Machine Learning-ICANN 2021: 30th International Conference on Artificial Neural Networks, Bratislava, Slovakia, September 14–17, 2021, Proceedings, Part V 30*. Springer, 2021, pp. 460–472.
- [24] D. Pang, L. H. Lin, and N. A. Smith, "Improving nat-

- ural language inference with a pretrained parser,” *arXiv preprint arXiv:1909.08217*, 2019.
- [25] M. A. S. Cabezudo, M. Inácio, A. C. Rodrigues, E. Casanova, and R. F. de Sousa, “Natural language inference for portuguese using bert and multilingual information,” in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2020, pp. 346–356.
- [26] E. Fonseca, L. Santos, M. Criscuolo, and S. Aluisio, “Assin: Avaliacao de similaridade semantica e inferencia textual,” in *Computational Processing of the Portuguese Language-12th International Conference, Tomar, Portugal*, 2016, pp. 13–15.
- [27] S. Wehnert, S. Dureja, L. Kutty, V. Sudhi, and E. W. De Luca, “Applying bert embeddings to predict legal textual entailment,” *The Review of Socionetwork Strategies*, vol. 16, no. 1, pp. 197–219, 2022.
- [28] M. Shajalal, M. Atabuzzaman, M. B. Baby, M. R. Karim, and A. Boden, “Textual entailment recognition with semantic features from empirical text representation,” in *International Conference on Speech and Language Technologies for Low-resource Languages*. Springer, 2022, pp. 183–195.
- [29] M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini, and R. Zamparelli, “Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment,” in *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, 2014, pp. 1–8.
- [30] N. Jiang and M.-C. de Marneffe, “Evaluating bert for natural language inference: A case study on the commitmentbank,” in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, 2019, pp. 6086–6091.

An Approach of Test Case Generation with Software Requirement Ontology

Adisak Intana, Kuljaree Tantayakul, Kanjana Laosen, Suraiya Charoenreh
College of Computing, Prince of Songkla University, Phuket, Thailand

Abstract—Software testing plays an essential role in software development process since it helps to ensure that the developed software product is free from errors and meets the defined specifications before the delivery. As the software specification is mostly written in the form of natural language, this may lead to the ambiguity and misunderstanding by software developers and results in the incorrect test cases to be generated from this unclear specification. Therefore, to solve this problem, this paper presents a novel hybrid approach, Software Requirement Ontologies based Test Case Generation (*ReqOntoTestGen*) to enhance the reliability of existing software testing techniques. This approach enables a framework that combines ontology engineering with the software test case generation approaches. Controlled Natural Language (CNL) provided by the ROO (Rabbit to OWL Ontologies Authoring) tools is used by the framework to build the software requirement ontology from unstructured functional requirements. This eliminates the inconsistency and ambiguity of requirements before test case generation. The OWL ontology resulted from ontology engineering is then transformed into the XML file of data dictionary. Combination of Equivalence and Classification Tree Method (CCTM) is used to generate test cases from this XML file with the decision tree. This allows us to reduce redundancy of test cases and increase testing coverage. The proposed approach is demonstrated with the developed prototype tool. The contribution of the tool is confirmed by the validation and evaluation result with two real case studies, Library Management System (LMS) and Kidney Failure Diagnosis (KFD) Subsystem, as we expected.

Keywords—Software testing; software requirement specification; ontology; test case; equivalence and classification tree method

I. INTRODUCTION

Software testing is one of the most important stages to detect errors in software development. The number of software bugs are not mainly caused by the code or design. One of the main causes of software bugs is from the specification [1][2]. As software specification gathered from the user's needs is mostly written in common natural languages in the Software Requirements Specification (SRS) document [3][4][5], this leads unstructured requirements to be ambiguous and misunderstood by software developers [4][6][7]. Furthermore, in system and user acceptance testing, test cases are generated from the SRS. This may result in incorrect test cases to be generated from the unclear specification. Therefore, it is necessary that the requirement specification needs to be very clear and well-defined before generating test cases.

Ontology engineering has been applied in Requirements Engineering (RE). An ontology is a formal representation of entities and relationships in a domain of interest [8]. As the semantics of concepts are formally defined, an ontology can be

used as a formal specification for a program. A domain vocabulary, essential concepts with their taxonomy, relationships (and constraints) between concepts, and domain axioms are defined for specific program applications [8][9]. Thus, using ontologies to express requirement specifications has implications for advantage in managing complexity, contradictions, or detecting ambiguity and incompleteness of requirements [4][10][11]. The application of ontology to requirement specification can help to eliminate the problem of erroneous test case generation from ambiguous, inconsistent, or incomplete requirements. Thus, our challenge is to add value to software testing with ontology modelling in requirement specification [12].

Therefore, in our previous work [13], we presented how ontology engineering approach can enhance practical software testing. We proposed a conceptual vision of framework called *ReqOntoTestGen* (Requirement Ontology Testcase Generation) that combines the benefit of ontology to represent the semantics of requirement specification with Control Natural Language (CNL) and Classification Tree Method (CCTM) [14][15] testing technique to generate test cases. The ROO (Rabbit to OWL Ontology Authoring) tool [16] is used by this framework to design and develop an ontology with CNL or Rabbit Language. This results in the complexity of requirements in natural languages to be reduced and the semantic of requirements formally defined. The specific syntax of this tool increases the structure and eliminates the ambiguity of the requirement ontology. The result of this tool is an export in Web Ontology Languages (OWL) format to transform into a structured data dictionary, before it is considered with decision tree to generate all possible test cases. Furthermore, CCTM provided by *ReqOntoTestGen* framework also allows the number of generated test cases to be minimized by reducing the redundant test cases and the testing coverage that covers all possible testing scenarios to be maximized. We demonstrated manually the effectiveness of the framework with a real case study, Library Management System (LMS).

The work of this paper is extended from the previous work [13]. This paper proposed a semi-automatic approach for test case generation from the requirement specification ontology based on use case-based requirement specification. To demonstrate the practical implementation of the approach, we developed a prototype tool according to *ReqOntoTestGen* framework in which the ontology engineering and test case generation algorithm is implemented in the tool. Control Natural Language (CNL) enabled by the ROO tool is used to be a guideline and build conceptual ontologies from the requirement specification. To generate test cases, the result from the ROO tool, the ontology represented in terms of OWL format, is transformed into the XML file of data dictionary. The OWL and XML transformation rules were designed and

implemented into our prototype tool for this data dictionary transformation. CCTM techniques were implemented in the tool for automatic test case generation purposes. The XML file of decision tree specifying the constraint of test case generation is considered with the XML file of data dictionary to generate test cases. Moreover, the validity, effectiveness and accuracy of the approach and tool were guaranteed by two different case studies formulated from real-world systems, Library Management System (LMS) and Kidney Failure Diagnosis (KFD) Subsystem. We compared the actual result of test case generation from these case studies by the tool with the expected test cases calculated manually by the practical testers. Furthermore, the satisfaction level of the proposed approach and tool was evaluated by practical specialists for future use.

The remainder of the paper is organised as follows. Firstly, Section II explains an overview of the necessary background and related work, before the proposed approach and the real-world case studies for experimenting the effectiveness of the approach are described in Section III. In Section IV, the proof of concept of our proposed approach are demonstrated through the evaluation of the implemented prototype with the case studies. Section V discusses the lesson learned experienced from the study result. Finally, the conclusion and future work are described in Section VI.

II. RELATED WORK

Several research studies have been interested in using ontology in the software development process to increase the efficiency of developed products. For instance, [17] proposed the software process automation ontology (Sponto) that applied the ontology-based approach to generate a set of artefacts for the software development process such as user stories for requirement specification and SQL for database scripts. Another example is the work of [18] which introduced the mechanism for transforming security requirements described in the form of the natural language into a structured ontology. The inconsistencies of security requirements were also checked by this mechanism.

However, most studies focus on using ontology to represent the conceptualisation and knowledge information regarding software development domains. [19], for example, proposed the application of ontology to define the information and knowledge semantics in RE. Instead of using ontology to represent the semantic of the requirement itself, this work focused on the use of ontology to describe the way of structuring requirements in the SRS document. Similar to this, [20], [21] and [22] proposed a domain ontology for software requirement change management, requirement classification and use story assessment in requirement artefacts respectively. In [23], they proposed ROoST (Reference Ontology on Software Testing) that builds a set of interrelated ontology patterns related to the software testing concepts including its process, activities, artefacts and testing techniques for test case design in order to associate semantics to a large amount of test information. Similar to this [24], [25] and [26] applied the ontology-based method to represent the knowledge related to software testing activities. The common well-established vocabulary for testing is used in the ontology application. Their developed ontologies influence the benefit of knowledge sharing among the development team.

Furthermore, some studies focus more on the application of requirement ontologies to generate test cases in practice. [23], for instance, presented a combined inference to software requirement ontology to generate test cases based on software requirement specification. The test cases were obtained from test input, test procedure, and expected test results. The work proposed by [23][27] used inference rules based on reasoner to generate test cases and improve requirements coverage and domain coverage. Furthermore, [28] presented Web Ontology Language for Web Service (OWL-S) to describe the workflow in the web service application. Petri-Net is used to represent the meaning of the test process and OWL-S is used to generate test data. In addition, [29] presented application of OWL ontologies to generate test cases and test procedures based on controlled-English model. The closely related work is proposed by [30]. They proposed test case generation using a learning-based software testing approach based on requirement ontology to generate test cases. However, those research studies mentioned earlier only focus on the application of requirement specification ontology to generate test cases, they did not consider testing coverage in test case generation. Based on our literature reviews, it can conclude that most of the existing research studies focus on using ontology to represent the software testing concept and knowledge sharing in software engineering communities. A few studies considered more important in the use of requirement specification ontology in the software testing process to generate comprehensive test cases together with testing coverage analysis of test case generation.

III. MATERIALS AND APPROACH

A. ReqOntoTestGen Framework

Fig. 1 shows a framework of the test case generation with software requirement ontology (*ReqOntoTestGen*) proposed in our previous work [13]. There are four steps in this framework. (1) *Ontology Engineering* generates the ontology according to CNL from the functional requirement definition described in terms of natural language by using ROO-CNL authoring. CNL in ROO authoring enables the complex requirement to be transformed into a very simple requirement before generating the ontology. Then, the achieved requirement ontology is exported in terms of OWL format, before (2) *XML Generation* transforms the exported OWL into the XML data dictionary metadata. In (3) *Variable and Decision Tree Management*, it starts with the variable information extracted from the XML of use case defined in the SRS document, before the corresponding data structure of the extracted variable is extracted from the XML data dictionary. This, then, is considered with the XML file of decision tree for test cases generation. Finally, (4) *Test Case Generation* creates test cases from the variable and its conditions by using CCTM technique. CCTM test case generation technique was chosen to be implemented in the proposed framework as it provides the benefit in which the number of test cases are minimized by eliminating the redundant test cases and the testing coverage is maximized in which all possible range value of test input variables is expanded.

B. ReqOntoTestGen Algorithm

To achieve a better understanding of our *ReqOntoTestGen* Framework explained in Section III-A, this section describes

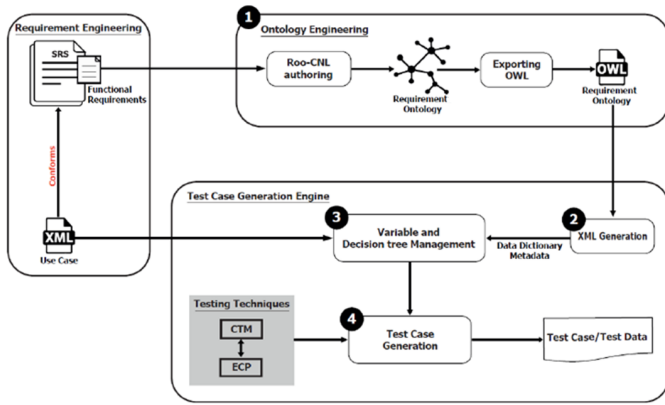


Fig. 1. The ReqOntoTestGen framework [13].

the algorithm of the framework in detail.

1) *Step 1: Ontology engineering:* In this step, the complexity of natural language based functional requirement and its corresponding constraints are reduced by transforming them into CNL structure. Then, the target ontology is developed from the transformed CNL based requirement. Table I shows an example of mapping from natural language-based requirement to ROO-CNL Structure and OWL2 Functional Syntax respectively. When classifying and structuring the ROO-CNL successfully, the ontology is exported in terms of file OWL format for data dictionary metadata generation in the next step.

TABLE I. EXAMPLE OF ROO STRUCTURE

Description	Roo-CNL Structure	OWL 2 Functional Syntax
Class Declaration	<i>cname</i> is a concept	Declaration(Class(: <i>cname</i>))
Subclass	Every <i>scname</i> is a kind of <i>cname</i>	SubClassOf(: <i>scname</i> : <i>cname</i>)
Relationship Declaration	<i>rname</i> is a relationship Every <i>cname1</i> <i>rname</i> <i>cname2</i>	Declaration(ObjectProperty(: <i>rname</i>)) ObjectPropertyDomain(: <i>rname</i> : <i>cname1</i>) ObjectPropertyRange(: <i>rname</i> : <i>cname2</i>)
Instance Declaration	<i>insname</i> is a <i>cname</i>	Declaration(NameIndividual(: <i>insname</i>)) ClassAssertion(: <i>cname</i> : <i>insname</i>)

2) *Step 2: The XML generation:* In this step, the OWL files obtained from the ontology are transformed into XML structures of data dictionary that is used for test case generation. The XML format is used for the target file transformed from the source of OWL file to make it easier to exchange data between programs [31]. Based on the study of [32][33], 13 relevant transformation rules are designed and used for transformation. All rules are available on our tool website¹.

Fig. 2 shows an example of transformation rules, consisting of the first column (Rules) as the rules of transformation. The second column (OWL2 Functional) is an OWL syntax. The last column (XML Schema) is an XML syntax. The transformation consists of three main categories, the structure of classes and relations, object property restrictions, and data property restrictions.

3) *Step 3: The variable and decision tree management:* This step considers two input files, the XML file of use

Rules	OWL2 Functional	XML Schema
The Structure of Classes and Relations		
1. Class Declaration	Class(: <i>cname</i>)	Local: <xs:element name=" <i>cname</i> "> <xs:complexType> </xs:complexType> </xs:element> Global: <xs:element name=" <i>cname</i> " />
2. Subclass Declaration	SubClassOf(: <i>scname</i> : <i>cname</i>)	<xs:element name=" <i>scname</i> " type=" <i>cname</i> " />
3. Object Property Declaration	ObjectProperty(: <i>hasClass</i>) ObjectPropertyDomain(: <i>hasClass</i> : <i>cdomain</i>) ObjectPropertyRange(: <i>hasClass</i> : <i>crange</i>)	<xs:element name=" <i>cdomain</i> "> <xs:complexType> <xs:sequence> <xs:element name=" <i>hasClass</i> " ref=" <i>crange</i> " /> </xs:sequence> </xs:complexType> </xs:element> <xs:element name=" <i>crange</i> "> <xs:complexType> </xs:complexType> </xs:element>
Object Property Restrictions		
5. Existential	EquivalentClasses(: <i>cdomain</i> ObjectSomeValuesFrom(: <i>hasClass</i> : <i>crange</i>))	<xs:element name=" <i>cdomain</i> "> <xs:complexType> <xs:sequence> <xs:element name=" <i>hasClass</i> " ref=" <i>crange</i> " /> </xs:sequence> </xs:complexType> </xs:element> <xs:element name=" <i>crange</i> "> <xs:complexType> <xs:choice> </xs:choice> </xs:complexType> </xs:element>
6. Universal	EquivalentClasses(: <i>cdomain</i> ObjectAllValuesFrom(: <i>hasClass</i> : <i>crange</i>))	<xs:element name=" <i>cdomain</i> "> <xs:complexType> <xs:sequence> <xs:element name=" <i>hasClass</i> " ref=" <i>crange</i> " /> </xs:sequence> </xs:complexType> </xs:element> <xs:element name=" <i>crange</i> "> <xs:complexType> <xs:all> <xs:all> </xs:complexType> </xs:element>
Data Property Restrictions		
11. Individual value	NameIndividual(: <i>individual1</i>) NameIndividual(: <i>individual2</i>) NameIndividual(: <i>individual3</i>) ClassAssertion(<i>cname1</i> : <i>individual1</i>) DataPropertyAssertion(: <i>dprop</i> : <i>individual1</i> "value1" <i>xsd:dtype</i>) ClassAssertion(<i>cname1</i> : <i>individual2</i>) DataPropertyAssertion(: <i>dprop</i> : <i>individual2</i> "value2" <i>xsd:dtype</i>) ClassAssertion(<i>cname1</i> : <i>individual3</i>) DataPropertyAssertion(: <i>dprop</i> : <i>individual3</i> "value3" <i>xsd:dtype</i>)	<xs:element name=" <i>dprop</i> "> <xs:complexType> <xs:choice> <xs:element name=" <i>value1</i> " type=" <i>xsd:dtype</i> " /> <xs:element name=" <i>value2</i> " type=" <i>xsd:dtype</i> " /> <xs:element name=" <i>value3</i> " type=" <i>xsd:dtype</i> " /> </xs:choice> </xs:complexType> </xs:element>
12. Minimum cardinality	EquivalentClasses(: <i>cname</i> DataMinCardinality(<i>min</i> : <i>dprop</i> <i>xsd:dtype</i>))	<xs:element name=" <i>cname</i> "> <xs:complexType> <xs:sequence> <xs:element name=" <i>dprop</i> " type=" <i>xsd:dtype</i> " minOccurs=" <i>min</i> " /> </xs:sequence> </xs:complexType> </xs:element>
13. Maximum cardinality	EquivalentClasses(: <i>cname</i> DataMaxCardinality(<i>max</i> : <i>dprop</i> <i>xsd:dtype</i>))	<xs:element name=" <i>cname</i> "> <xs:complexType> <xs:sequence> <xs:element name=" <i>dprop</i> " type=" <i>xsd:dtype</i> " maxOccurs=" <i>max</i> " /> </xs:sequence> </xs:complexType> </xs:element>

Fig. 2. Example of OWL and XML transformation rules.

cases and the XML file of data dictionary transformed from the OWL of requirement ontology. The use case files are designed from requirements in the SRS document according to UML Development Guidelines Version 2.0 [34]. The use case normally demonstrates the overview of functionality and procedure of the system to generate test cases. Fig. 3 shows

¹ <https://sites.google.com/phuket.psu.ac.th/reqontotestgen/>

an example of brief description of use case UC001 describing the behaviour of function *Borrow Item* of LMS. To represent the function behaviour, a sequence of the step-by-step is used including main flow of events for the most common success scenario, alternative flow of events for other less common success scenario and exception flow of events for the error management scenario. The input and output for function operation, then, are indicated from the use case corresponding steps. The data structure of input and output variables are described in the XML files of data dictionary transformed from the OWL of requirement ontology.

Use Case	Description
UC001-Borrow Item	<p>Use Case ID: UC001</p> <p>Use Case Name: Borrow Item</p> <p>Pre-Condition: Checking members of the library</p> <p>Post-Condition: Borrow items successfully</p> <p>Priority: High</p> <p>Flow of Event: 1. The system shows GUI for a list of items to borrow 2. Member select items to borrow 3. Member confirm borrow items [A1] 4. The system records the borrowed items 5. The system displays a list of borrowed items</p> <p>Alternative Flow: [A1] If click "Cancel" button, the system will not records [E1]</p> <p>Exception Flow: [E1] The system shows the warning "Confirm cancellation?"</p>

Fig. 3. Example of use case detail.

4) *Step 4: Test case generation:* Our framework implements CCTM technique for test case generation. It generates test cases from the extracted input variable with the corresponding data structure. The test case generation process is described as follows.

Step 4.1: Classification Tree Generation with CTM Technique. CTM technique generates a classification tree from the information extracted from the use case. It starts with the name of the system represented by the use case name to be a root node of the tree. Then, it layers the tree from the root node to the terminal classification node with the subsystem and its corresponding variables respectively. The leaf node of the tree, terminal class, defines the range of variable values which are considered to create partitions for both valid and invalid data values by ECP technique. This data range value is used to generate test cases in the later step. An example of classification tree for function *Borrow Item* of LMS resulted from CTM is shown in Fig. 4. The variables and their corresponding range value are visualised in terminal classification (parent node) and terminal class (leaf node) of the tree respectively. This can be explained as follows: $Member = \{AdminStaff, Grad, Lecturer, Undergrad, None\}$, $Item = \{Book, CD, DVD, None\}$, $borrowDate = \{beginDate-endDate, None\}$ and $maxDaysBorrow = \{7, 14, 30, None\}$. These are considered to create an equivalence class partitioning in the next step.

Step 4.2: Test Case Generation with ECP Technique. In the



Fig. 4. Example of a classification tree for function *Borrow Item* of LMS.

classification tree achieved from CTM technique, ECP divides the terminal classification into equivalence classes for each possible range of data values. The framework implemented a strong robust format [4] to generate a test case. The equivalence class in this form considers both valid and invalid values of all classes of equivalence and allows the test case generation to cover every possible value of all equivalence classes. An example ECP for function *Borrow Item* of LMS is shown in Fig. 5.

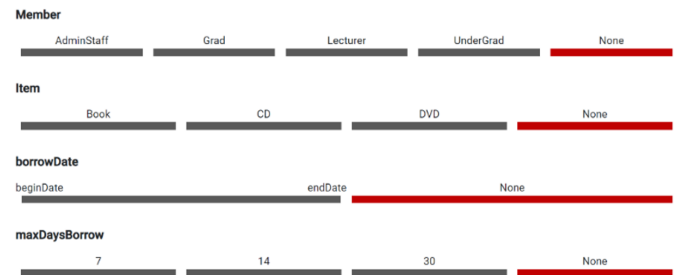


Fig. 5. Example of equivalence class partitioning for function *Borrow Item* of LMS.

C. Case Studies

To demonstrate the effectiveness of our proposed approach, case studies from the real world system are used. We consider two different case studies for this purpose. One is a Library Management System (LMS) deployed in Prince of Songkhla University, Phuket. The other is Kidney Failure Diagnosis (KFD) subsystem from Hospital Information System (HIS) replicated from [15]. The following sections describe the case study information together with the demonstration of how our approach manually works.

1) *Library Management System (LMS):* The LMS² is a system for managing various library resources. The members of the library can borrow or return resources such as books, CDs, or DVDs. Each type of member has different borrowing

²<http://library.phuket.psu.ac.th/>

conditions. To generate a test case, we considered the return function that contains fine calculation when the late return occurs. The detailed information as well as practical requirements were formulated from the LMS of Prince of Songkla University.

Requirements: The functional requirements of LMS is shown in Table II.

TABLE II. THE FUNCTIONAL REQUIREMENTS OF LMS

Req. ID	Requirements
LIB-FUN-01	Members can borrow item including books, CDs or DVDs.
LIB-FUN-02	Members are classified into admin staff, graduate student, lecturer and undergraduate.
LIB-CON-01	Books, CDs, DVDs must be disjointed.
LIB-CON-02	Admin, staff, graduate student, lecturer and undergrade must be disjointed.
LIB-CON-03	Maximum borrowing items and borrowing periods: 5 books per 7 days for admin staff and undergraduate students, 10 books per 14 days for graduate students, and 15 books per 30 days for lecturers.
LIB-CON-04	Maximum borrowing items and borrowing periods: 3 discs per 7 days for all members.

Ontology Engineering: From the functional requirements of LMS as shown in Table II, it can be used to design and develop an ontology which consists of classes, relationships, and data properties. The ontology syntax of LMS is shown in Fig. 6.

Req. ID	Roo-CNL Structure	OWL2 Functional Syntax
LIB-FUN-01	<i>Member</i> is a concept <i>Item</i> is a concept Every <i>Book</i> is a kind of <i>Item</i> Every <i>CD</i> is a kind of <i>Item</i> Every <i>DVD</i> is a kind of <i>Item</i> <i>hasBorrow</i> is a relationship Every <i>Member</i> <i>hasBorrow</i> <i>Item</i>	Declaration(Class(: <i>Member</i>)) Declaration(Class(: <i>Item</i>)) SubClassOf(: <i>Book</i> : <i>Item</i>) SubClassOf(: <i>CD</i> : <i>Item</i>) SubClassOf(: <i>DVD</i> : <i>Item</i>) Declaration(ObjectProperty(: <i>hasBorrow</i>)) ObjectPropertyDomain(: <i>hasBorrow</i> : <i>Member</i>) ObjectPropertyRange(: <i>hasBorrow</i> : <i>Item</i>)
LIB-FUN-02	Every <i>AdminStaff</i> is a kind of <i>Member</i> Every <i>Grad</i> is a kind of <i>Member</i> Every <i>Lecturer</i> is a kind of <i>Member</i> Every <i>UnderGrad</i> is a kind of <i>Member</i>	SubClassOf(: <i>AdminStaff</i> : <i>Member</i>) SubClassOf(: <i>Grad</i> : <i>Member</i>) SubClassOf(: <i>Lecturer</i> : <i>Member</i>) SubClassOf(: <i>UnderGrad</i> : <i>Member</i>)
LIB-CON-01	<i>DisjointClasses</i> ()	DisjointClasses(: <i>Book</i> : <i>CD</i>) DisjointClasses(: <i>Book</i> : <i>DVD</i>) DisjointClasses(: <i>CD</i> : <i>DVD</i>)
LIB-CON-02	<i>DisjointClasses</i> ()	DisjointClasses(: <i>AdminStaff</i> : <i>Grad</i>) DisjointClasses(: <i>AdminStaff</i> : <i>Lecturer</i>) DisjointClasses(: <i>AdminStaff</i> : <i>UnderGrad</i>) DisjointClasses(: <i>Grad</i> : <i>Lecturer</i>) DisjointClasses(: <i>Grad</i> : <i>UnderGrad</i>) DisjointClasses(: <i>Lecturer</i> : <i>UnderGrad</i>)
LIB-CON-03	<i>borrowDaysR1</i> is a <i>maxDaysBorrow</i>	Declaration(NameIndividual(: <i>borrowDaysR1</i>)) ClassAssertion(: <i>AdminStaff</i> : <i>borrowDaysR1</i>) DataPropertyAssertion(: <i>maxDaysBorrow</i> : <i>borrowDaysR1</i> "7"^^xsd:integer)
LIB-CON-04	And, configure data property assertion directly through GUI in the tool.	
LIB-FUN-03	<i>hasReturn</i> is a relationship Every <i>Member</i> <i>hasReturn</i> <i>Item</i>	Declaration(ObjectProperty(: <i>hasReturn</i>)) ObjectPropertyDomain(: <i>hasReturn</i> : <i>Member</i>) ObjectPropertyRange(: <i>hasReturn</i> : <i>Item</i>)
LIB-FUN-04	<i>DataProperty</i> () <i>DataPropertyDomain</i> () <i>ObjectSomeValuesFrom</i> () <i>DataPropertyRange</i> ()	Declaration(DataProperty(: <i>fine</i>)) DataPropertyDomain(: <i>fine</i> : <i>Member</i>) ObjectAllValuesFrom(: <i>hasReturn</i> : <i>Member</i>) DataPropertyRange(: <i>fine</i> :xsd:integer)

Fig. 6. The ontology syntax of LMS.

Fig. 7 shows the ontology structure of LMS generated by ROO tool. It consists of two classes that are related to each other. The *Member* class is a member of the library including *AdminStaff*, *Grad*, *Lecturer*, and *UnderGrad*. The *Item* class is a library resource that can be borrowed including *Book*, *CD*, and *DVD*. The *ObjectProperty* between the *Member* and *Item* classes represents the relationship in which members can borrow (*hasBorrow*) library resources. Another relationship, *hasReturn* is a relationship where members can return library resources after they have been borrowed. Furthermore, the *DataProperty* is also an entity of data, the domain is a

class, and the data type is a range of data properties. For example, *borrowDate* has class *Member* to be a domain and *xsd:dateTime* to be a range. Moreover, an individual or instance of value such as *borrowDayR1* "7" is the condition for the maximum of days to borrow the *Book* of *UnderGrad* member type.

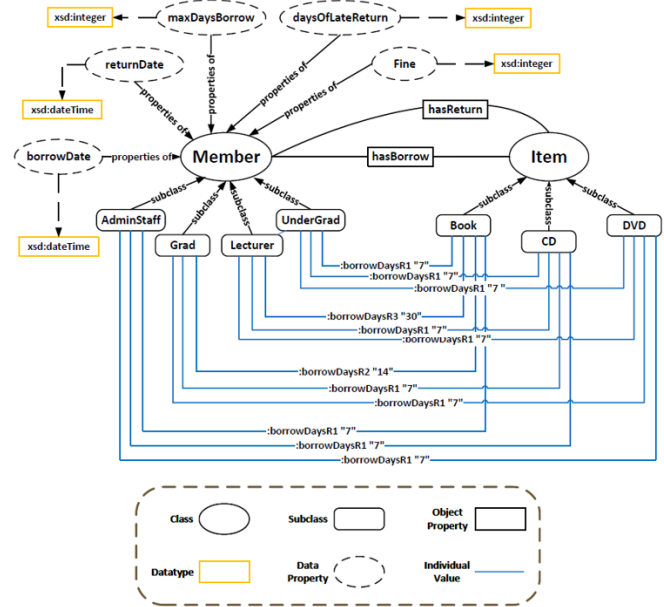


Fig. 7. The ontology structure of LMS.

Test Case Generation: Fig. 8 demonstrates the classification tree resulted from CTM technique. In the tree, *Library Management System* as a system name is considered to be a root node, before the subsystem *Return Item* is defined as a terminal classification in the next level. Variables *Member*, *Item*, *borrowDate*, *returnDate*, *maxDaysBorrow*, *daysOfLateReturn*, and *fine* related to this function are defined in the below level in the tree. These variables are considered to generate test cases by using ECP in the later step. The terminal class of each terminal classification defining the possible range of value is used to be a partition for generating test cases and test data in ECP.

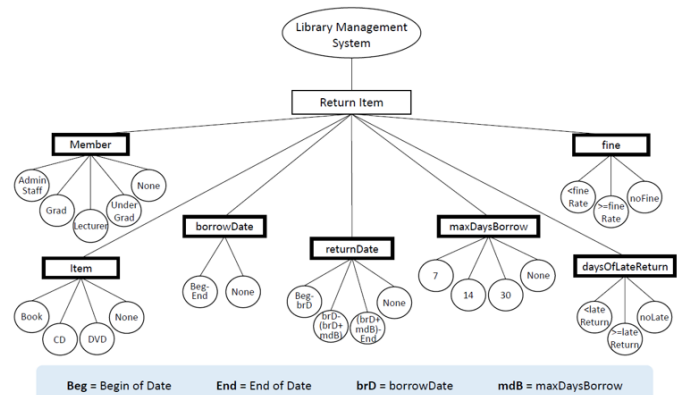


Fig. 8. The classification tree of LMS.

Test cases are generated from the Cartesian product of all

equivalence classes defined in four input variables (*Member*, *Item*, *borrowDate*, *returnDate*). There are a total of 160 (5*4*2*4) test cases to be generated. An example of test cases and test data is shown in Table III.

TABLE III. EXAMPLE OF TEST CASES AND TEST DATA FOR TESTING RETURN OPERATION

TC#	Member	Item	borrow Date	return Date	maxDays Borrow	daysOf LateReturn	fine	Comments
1	AdminStaff	Book	1/7/2019	2/7/2019	7	0	0	Valid
2	AdminStaff	Book	5/7/2019	15/7/2019	7	3	9	Valid
...
67	Lecturer	Book	10/7/2019	9/6/2019	30	-1	-3	Invalid
68	Lecturer	Book	10/7/2019	None	30	None	None	Invalid
...
160	None	None	None	None	None	None	None	Invalid

Table III is the test case generation result for testing return operation. Consider test case #1, it is a normal test case in which there is no late return for *AdminStaff*. This test case is different from test case #2. This results in the fine of 9 (3*3) to be calculated. Furthermore, the generated test cases cover in the case of invalid. In invalid test case #67, for example, it defines the return date before the borrowing date.

2) *Kidney Failure Diagnosis (KFD) Subsystem*: The KFD subsystem is a system for recommending the treatment appropriately to physicians for patients that have kidney dysfunction. It is calculated from the *Glomerular Filtration Rate (GFR)* result, consisting of *sex*, *age*, and *creatinine result (SCr)*. The *GFR* and *Urine Creatinine (UO)* results are paired to interpret the stage of kidney failure. This case study is based on [15]. It is an open-source system and is part of the Hospital Information System called HospitalOS³. It is a system that is installed and used in community hospitals and more than 100 clinics in Thailand.

Requirements: The functional requirements of KFD that design and develop an ontology comprise a total of four requirements as shown in Table IV.

TABLE IV. THE FUNCTIONAL REQUIREMENTS OF KFD

Req. ID	Requirements
KFD-FUN-01	Stage is paired with GFR and UO.
KFD-FUN-02	Stage of GFR includes ESRD, Loss, Failure, Injury and Risk.
KFD-CON-01	ESRD, Loss, Failure, Injury, Risk must be disjointed.
KFD-CON-02	GFR is calculated with sex, age, height and SCr.

Ontology Engineering: From the functional requirements of KFD in Table IV, it can be used to design and develop an ontology which consists of classes, relationships, and data properties. The ontology syntax of KFD is shown in Fig. 9.

Fig. 10 shows the ontology structure of KFD resulted from ROO tool. It consists of three classes: *Stage*, *GFR*, and *UO*. The stage of kidney failure includes *ESRD*, *Loss*, *Failure*, *Injury*, and *Risk*. The *ObjectProperty* is the relationship between classes. For example, *hasPair* is a relationship between *GFR* and *UO* class to represent a pair to interpret the stage of kidney failure. Furthermore, the *DataProperty* is also an entity of data. As *GFR* contains *Scr*, *Height*, *Age* and *Sex*, they are defined as a data property. In the data property, the domain is a class and the data type is a range. For example, *Height* has class *GFR*

Req. ID	Roo-CNL Structure	OWL2 Functional Syntax
KID-FUN-01	Stage is a concept GFR is a concept UO is a concept hasGFR is a relationship Every Stage hasGFR GFR	Declaration(Class(:Stage)) Declaration(Class(:GFR)) Declaration(Class(:UO)) Declaration(ObjectProperty(:hasGFR)) ObjectPropertyDomain(:hasGFR :Stage) ObjectPropertyRange(:hasGFR :GFR)
KID-FUN-02	Every ESRD is a kind of Stage Every Loss is a kind of Stage Every Failure is a kind of Stage Every Injury is a kind of Stage Every Risk is a kind of Stage	SubClassOf(:ESRD :Stage) SubClassOf(:Loss :Stage) SubClassOf(:Failure :Stage) SubClassOf(:Injury :Stage) SubClassOf(:Risk :Stage)
KID-CON-01	DisjointClasses()	DisjointClasses(:ESRD :Loss) DisjointClasses(:ESRD :Failure) DisjointClasses(:ESRD :Injury) DisjointClasses(:ESRD :Risk) DisjointClasses(:Loss :Failure) DisjointClasses(:Loss :Injury) DisjointClasses(:Loss :Risk) DisjointClasses(:Failure :Injury) DisjointClasses(:Failure :Risk) DisjointClasses(:Injury :Risk)
KID-FUN-03	DataProperty() DataPropertyDomain() DataPropertyRange() sex1 is a Sex And. configure data property assertion directly through GUI in the tool.	Declaration(DataProperty(:Sex)) DataPropertyDomain(:Sex :GFR) DataPropertyRange(:Sex xsd:integer) Declaration(NameIndividual(:sex1)) DataPropertyAssertion(:Sex :sex1 "Female"^^xsd:string)

Fig. 9. The ontology syntax of KFD.

to be a domain and *xsd:integer* to be a class range including the restriction of data property is 0-300 (0-300^^xsd:integer). Another example is *Sex* which has a domain to be class *GFR* and a range to be *xsd:string*. For this property, two individuals or instances are defined *Female* and *Male* to represent the gender of the patient.

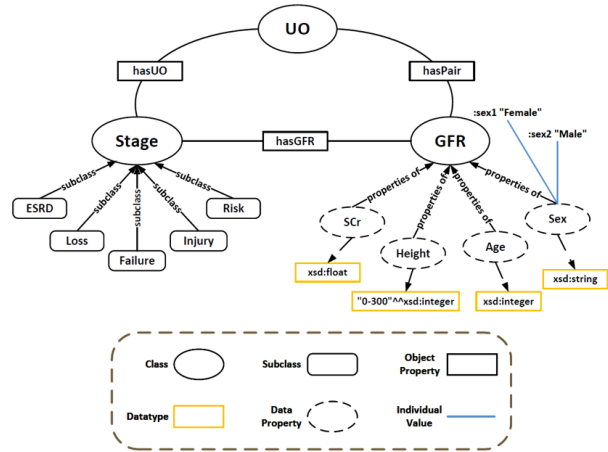


Fig. 10. The ontology structure of KFD.

Test Case Generation: Fig. 11 demonstrates the classification tree resulted from CTM technique. In the tree, *GFR Module* as a system name is considered to be a root node, before the subsystem *GFR Interpreted* is defined as a terminal classification in the next level. Variables *Sex*, *Age*, *Height*, *Scr*, *GFR*, *UO*, and *Stage* related to this function are defined in the below level in the tree. These variables are considered to generate test cases by using ECP in the later step. The terminal class of each terminal classification defining the possible range of value is used to be a partition for generating test cases and test data in ECP.

Table V is the test case generation result for testing *GFR interpreted* operation. Consider test case #1, it is a valid test case for *GFR* calculation of a female patient less than 18 years old. This result of the stage of kidney failure interpreted as the *Injury*. Furthermore, the generated test cases cover in the case

³http://www.opensource-technology.com

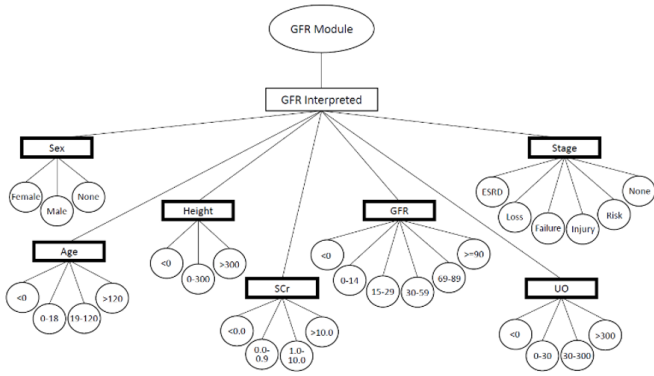


Fig. 11. The classification tree of KFD.

of invalid. In test case #144, it is an invalid test case because the data value is out of the range of interest.

TABLE V. EXAMPLE OF TEST CASES AND TEST DATA FOR TESTING RETURN OPERATION

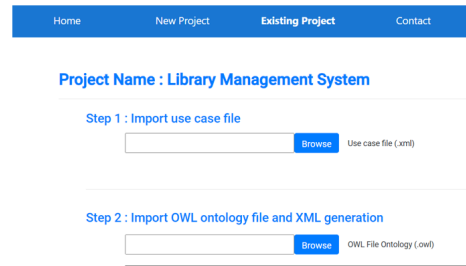
TC#	Sex	Age	Height	Scr	GFR	UO	Stage	Comments
1	Female	10	142	0.8	73	450	Injury	Valid
2	Female	18	165	4.5	15	27	Loss	Valid
...
61	Male	177	0.3	356	30	555	Risk	Valid
62	Male	65	104	7.2	7	10	ESRD	Valid
...
144	None	200	1140	110.2	11558	65487	None	Invalid

IV. PROOF OF CONCEPT

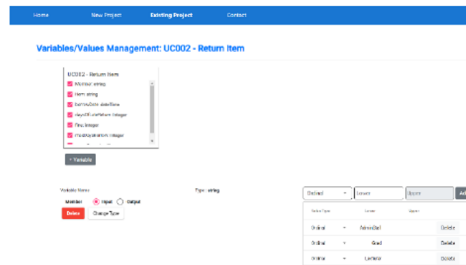
A. Tool Development

To demonstrate the effectiveness of *ReqOntoTesGen* approach, a prototype tool was developed. The developed tool is a Java based web application using Node.js 16.14.0⁴ JavaScript runtime environment that is well known and widely used.

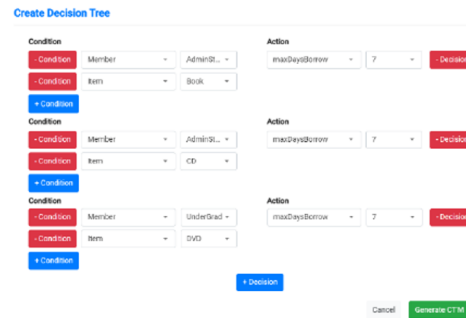
Fig. 12 demonstrates an example of our developed tool. Fig. 12a shows the screen for importing the necessary XML file. Two types of XML files are imported into the tool (1) the XML file of use cases indicating functionality from SRS documents and (2) the OWL file of requirement specifications created by the ROO tool. Then, the XML file of data dictionary is automatically generated from the OWL file. All extracted variables and their range value from the XML file of data dictionary are analysed. This includes variable name, variable type and variable range value as shown in Fig. 12b. The next step is the decision tree creation in the case that the system uses the condition for decision making on the operation process. The condition and decision of the decision tree can be adjusted as necessary as demonstrated in Fig. 12c. This decision tree is considered with the transformed data dictionary to generate test cases by the CCTM technique in the tool. The classification tree and equivalence partition of related variables resulted from CCTM are shown on the screen as demonstrated in Fig. 4 and 5 respectively. Test cases are automatically generated from this classification tree and equivalence partition. The result of test case generation is shown on the screen as in Fig. 12d.



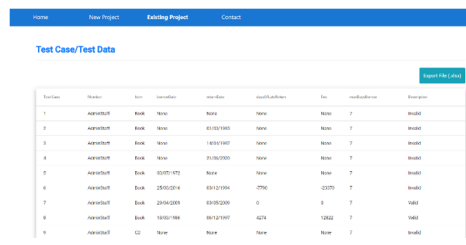
(a) The import file management screen



(b) The variables and values management screen



(c) The screen for adding decision tree



(d) The test case / test data generation screen

Fig. 12. Example of the tool screens.

B. Tool Validation

To validate the developed tool whether all functionalities of the tool perform correctly according to the *ReqOntoTestGen* framework proposed in Section III. Three test scenarios corresponding to three steps of the framework were conducted as shown in Table VI. This includes 1) *TS-01 Validate OWL transformation to XML function* with 13 relevant designed transformation rules. 2) *TS-02 Validate variable and decision tree management function* to validate the correctness of extracted variables from the XML file of use case and data dictionary together with the decision tree information. 3) *TS-03 Validate test case generation function* that validates the correctness of test case generation with CCTM techniques.

⁴<https://nodejs.org/en/about/>

TABLE VI. THE RESULT OF TOOL TESTING

Test Scenario	Result	Revision
TS-01 Validate OWL transformation to XML function	Fail	Pass
TS-02 Validate variable and decision tree management function	Pass	-
TS-03 Validate test case generation function	Pass	-

Table VI demonstrates the validation result. This led us to reveal an error that occurred in *TS-01*. The validation result of *TS-01* was *Fail* because the individual value transformation rule generated the wrong order in the XML element as shown in Fig. 13a This resulted in the data property element e.g., *maxDaysBorrow* to be generated outside the class element. This violated our designed transformation rules in which the data property needs to be inside the class. This led us to restructure the rule according to the design. Fig. 13b shows the corrected version of this error that resulted in this testing scenario to be *Pass*.

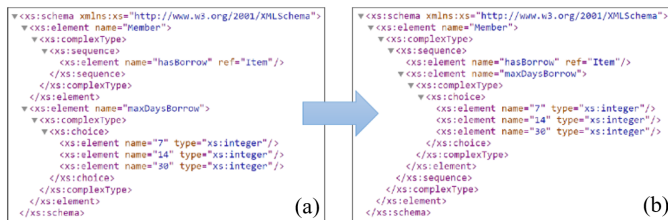


Fig. 13. The error of *TS-01*.

C. Tool Evaluation

We evaluated the effectiveness of our proposed approach with two real case studies, Library Management System (LMS) and Kidney Failure Diagnosis (KFD) subsystem. This evaluation is divided into two parts that are 1) effectiveness evaluation and 2) satisfaction evaluation.

1) *Effectiveness evaluation*: The precision, recall and F-measure computation were calculated by comparing the result produced by the manual operation and automated tool. The computation metrics were adapted from [35] as follows.

$$Precision = \frac{|\{Expert\ Identified\} \cap \{Tool\ Identified\}|}{|\{Tool\ Identified\}|} \times 100 \quad (1)$$

$$Recall = \frac{|\{Expert\ Identified\} \cap \{Tool\ Identified\}|}{|\{Expert\ Identified\}|} \times 100 \quad (2)$$

$$F - measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

TABLE VII. THE RESULT OF IMPACT ANALYSIS

System	# Test cases		Precision	Recall	F-measure
	identified by an expert	identified by the tool			
LMS	300	160	100%	53.33%	69.56%
KFD	144	144	100%	100%	100%

Table VII demonstrates the comparison results between the expected test case manually created by experts and the actual test case automatically generated by the tool. Considering the calculated F-measure with precision and recall of KFD case study, the accuracy of the automatic tool performing with this

case study is very high. This is because KFD case study is not a complex case study compared to LMS case study. However, considering the calculated F-measure, with precision and recall of LMS, they are quite low. We have found that in the manual design of test cases by experts, the out of range of variable *borrowDate* and *returnDate* was identified as invalid partition. This led to 300 (5*4*3*5) test cases to be created. However, after we revealed this case, we discovered that this type of *dateTime* variable has the range of time from “*Begin of Date*” to “*End of Date*” that can be selected at any time for testing. Therefore, it is impossible to be “*Out of Range*”. This led us to recalculate the number of created test case after cutting these “*Out of Range*” partition (partitions 12 and 13 in Fig. 14) and resulted in this recalculation to be the same as calculated by the tool.

Partition	Num of Partition	Valid Data	Invalid Data
Member (input)	5	AdminStaff ⁽¹³⁾ , Grad ⁽¹⁴⁾ , Lecturer ⁽¹⁵⁾ , UnderGrad ⁽¹⁶⁾	None ⁽⁵⁾
Item (input)	4	Book ⁽¹⁶⁾ , CD ⁽¹⁷⁾ , DVD ⁽¹⁸⁾	None ⁽⁹⁾
borrowDate (input)	3	Begin of Date – End of Date ⁽¹⁹⁾	None ⁽¹¹⁾ , Out of Range ⁽¹²⁾
returnDate (input)	5	hasBorrowDate – ⁽¹³⁾ (hasBorrowDate + hasMaxDaysBook) – End of Date ⁽¹⁴⁾	Begin of Date – ⁽¹⁵⁾ (hasBorrowDate + hasBorrowDate) – End of Date ⁽¹⁶⁾ , None ⁽¹⁶⁾ , Out of Range ⁽¹⁷⁾
maxDaysBorrow (fix rate)	4	7 ⁽¹⁸⁾ , 14 ⁽¹⁹⁾ , 30 ⁽²⁰⁾	None ⁽²¹⁾
daysOfLateReturn (output)	3	1 – Late Return ⁽²²⁾ , None ⁽²³⁾	Smallest Number – 1 ⁽²⁴⁾
fine (output)	3	1 x Fine Rate – ⁽²⁵⁾ daysOfLateReturn x Fine Rate	None ⁽²⁶⁾ , Smallest number – Fine Rate ⁽²⁷⁾

Fig. 14. The partition of variable change.

2) *Satisfaction evaluation*: The satisfaction of our proposed approach and tool was evaluated with a wide range of experts that have at least five years in software engineering and software testing. This included two programmers and three testers. We designed questions for satisfactory evaluation, *Q1) Functionality*, *Q2) Efficiency and reliability*, *Q3) Usability*, and *Q4) Ability and applicability* that is shown in Table VIII.

TABLE VIII. SATISFACTION QUESTIONS

Questions	Average
Q1. Functionality	
Q1.1 The function can operate accurately and appropriately.	Likert scale (Mandatory)
Q1.2 The function can operate with each other.	Likert scale (Mandatory)
Q1.3 The function can operate according to the users' requirements.	Likert scale (Mandatory)
Q2. Efficiency and reliability	
Q2.1 The prototype can appropriately process the test cases.	Likert scale (Mandatory)
Q2.2 The prototype can increase the structure of functional requirements.	Likert scale (Mandatory)
Q2.3 The prototype can reduce errors caused by functional requirements.	Likert scale (Mandatory)
Q2.4 The prototype can work completely.	Likert scale (Mandatory)
Q3. Usability	
Q3.1 The prototype is easy to learn and understand.	Likert scale (Mandatory)
Q3.2 The prototype is easy to use, and the function is not complicated.	Likert scale (Mandatory)
Q4. Ability and applicability	
Q4.1 The prototype can be applied in the system or other case studies.	Likert scale (Mandatory)
Q4.2 The prototype can be easily installed and used.	Likert scale (Mandatory)

The Likert scale was used to design the levels of satisfactory for each question including Strongly Agree (5), Agree (4),

Neutral (3), Disagree (2), and Strong Disagree (1) respectively. The evaluation result of the satisfactory is shown in Fig. 15.

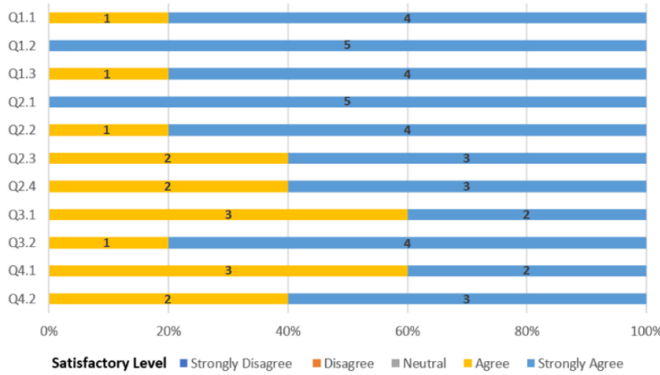


Fig. 15. Results of the four Likert scale questions.

As can be seen in Fig. 15, most of the specialists strongly agreed that our developed tool provides an accurate and appropriate functionality and interoperability (with an average of Q1.1-1.3, 86.66% of them strongly agree). They also strongly agreed that the tool is efficient and reliable (with an average of Q2.1-2.4, 75% of them strongly agree). Considering the usability (Q3), 60% of specialists strongly agreed that the functions provided by the tool are not complicated and easy to use and understand. Furthermore, the specialists satisfied the prototype in terms of its ability and applicability (Q4) from the agree level (with half of them satisfied at the strongly agree level). Overall, we can conclude that the specialists were mostly satisfied our *ReqOntoTestGen* approach and its corresponding tool.

V. LESSON LEARNED AND DISCUSSION

In this section, we discuss the benefits of the proposed *ReqOntoTestGen* approach for generating test cases with the software requirement ontology. This section also shares lessons learned achieved from our implications of practical implementation and experiment. The approach influences the benefits according to our research questions as follows.

- It provides a systematic mechanism and framework to generate test cases from a very clear structure of functional requirements encoded in the form of ontology. The application of ROO tool in the framework enables the unstructured requirement to be transformed into more structured and clearer requirements before generating ontology. This results in the complexity of requirement structure to be reduced and the ambiguity of the terminology used in the ontology to be eliminated as discussed in [8][11][16][36][37]. This also guarantees that the main causes of errors in software testing that are mainly from requirements to be eliminated and the correct test cases that satisfied user requirements to be generated.
- CCTM test case generation technique implemented in the framework to construct test cases influences benefits that the number of test cases is reduced with maximizing testing coverage. As claimed in [14][15] we have discovered from our implemented experiences

that CTM technique in CCTM enables the redundant test cases to be eliminated, On the other hand, ECP technique in CCTM expands the possible range value both valid and invalid cases. This led to the testing coverage to be increased.

- *ReqOntoTestGen* approach provides a semi-automatic prototype tool that implemented the algorithm to generate test cases from well-defined ontology. The results of the experiment by comparing the manual test case generation and automatic test case generation by the tool with two case studies: LMS and KFD can guarantee the correctness, effectiveness, and accuracy of the proposed approach and tool. Furthermore, the efficiency and potential use in the future are confirmed by the evaluation result from experts.

However, as suggested by the practical specialists from the satisfaction evaluation, there are limitations of the approach. Firstly, the proposed approach provides the semi-automated prototype tool in which the conceptual ontologies from the requirement specification resulted from the ROO tool need to be input manually into the prototype for test case generation. Furthermore, the experiment for the prototype validation and evaluation is based on two real case studies. It needs to be evaluated with other different domain of case studies.

VI. CONCLUSION AND FUTURE WORK

This paper presents a novel approach, *ReqOntoTestGen*, to enhance the efficiency of traditional testing techniques. It provides a semi-automatic framework that integrates ontology engineering with software testing for test case generation. The effectiveness and efficiency of our *ReqOntoTestGen* approach and framework is demonstrated by the developed prototype tool. The experiment results with the implementation of two case studies have shown that the Control Natural Language (CNL) from the ROO tool used in our tool enables the unstructured functional requirements that may lead the generated test cases to be inconsistent to the users' needs to be more structured and clearer, before transforming them into the OWL conceptual ontology. This OWL file is, then, transformed automatically into the XML file of data dictionary. CCTM technique implemented in the tool creates the automatic test case generation environment in which test cases are generated automatically from the transformed XML file of data dictionary with the decision tree. This influences the benefits that the redundant test cases to be eliminated and the coverage of the test case generation to be increased. Furthermore, the evaluation result has shown that our developed tool has a high degree of validity, accuracy and satisfaction level from the practical specialist perspective. As a result of this, it can be confirmed that our proposed approach contributes a hybrid test case generation technique with a software requirement ontology engineering that both meets the users' need and covers all possible testing scenarios.

For the future work, to increase the capability and reliability of the developed prototype, it needs to link with the ROO tool which can automatically input the conceptual ontology resulted from the ROO to the prototype. Furthermore, the evaluation of the prototype with different domain of case studies is still open as another research issue.

DEPLOYMENT AND AVAILABILITY

The developed tool with the user guide document and source of example case studies is available at <https://sites.google.com/phuket.psu.ac.th/reqontotestgen/>.

REFERENCES

- [1] R. Patton, *Software Testing (2nd Edition)*. USA: Sams, 2005.
- [2] Z. Liu and R. Kang, "Imperfect debugging software belief reliability growth model based on uncertain differential equation," *IEEE Transactions on Reliability*, vol. 71, no. 2, pp. 735–746, 2022.
- [3] D. Dermeval, J. Vilela, I. I. Bittencourt, J. Castro, S. Isotani, P. Brito, and A. Silva, "Applications of ontologies in requirements engineering: A systematic review of the literature," *Requirements Engineering*, vol. 21, no. 4, p. 405–437, nov 2016. [Online]. Available: <https://doi.org/10.1007/s00766-015-0222-6>
- [4] K. Thongglin, S. Cardey, and P. Greenfield, "Thai software requirements specification pattern," in *2013 IEEE 12th International Conference on Intelligent Software Methodologies, Tools and Techniques (SoMeT)*, 2013, pp. 179–184.
- [5] G. Liargkovas, A. Papadopoulou, Z. Kotti, and D. Spinellis, "Software engineering education knowledge versus industrial needs," *IEEE Transactions on Education*, vol. 65, no. 3, pp. 419–427, 2022.
- [6] P. Jorgensen, *Software Testing: A Craftsman's Approach*, 3rd ed. Boca Raton, NY: Auerbach Publications, 5 2013.
- [7] K. Mokos, T. Nestoridis, P. Katsaros, and N. Bassiliades, "Semantic modeling and analysis of natural language system requirements," *IEEE Access*, vol. 10, pp. 84 094–84 119, 2022.
- [8] C. Keet. (2020) An introduction to ontology engineering. [Online]. Available: <https://people.cs.uct.ac.za/~mkeet/OEbook/>
- [9] W. W. Sim and P. Brouse, "Towards an ontology-based persona-driven requirements and knowledge engineering," *Procedia Computer Science*, vol. 36, pp. 314–321, 2014, complex Adaptive Systems Philadelphia, PA November 3-5, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1877050914013489>
- [10] K. Siegemund, U. Assmann, J. Pan, E. Thomas, and Y. Zhao, "Towards ontology-driven requirements engineering," in *Proceeding of the 10th International Semantic Web Conference (ISWC)*, 10 2011.
- [11] N. S. Harsha, C. N. Kumar, V. K. Sonthi, and K. Amarendra, "Lexical ambiguity in natural language processing applications," in *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, 2022, pp. 1550–1555.
- [12] S. Popereshnyak and A. Vecherkovskaya, "Modeling ontologies in software testing," in *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, vol. 3, 2019, pp. 236–239.
- [13] S. Charoenreh and A. Intana, "Enhancing software testing with ontology engineering approach," in *2019 23rd International Computer Science and Engineering Conference (ICSEC)*, 2019, pp. 186–191.
- [14] B. Ramadoss, P. Prema, and S. R. Balasundaram, "Combined classification tree method for test suite reduction," in *Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET, 2011)*, no. 11, 2011, pp. 27–33.
- [15] A. Intana, K. Laosen, and T. Sriraksa, "An automated impact analysis approach for test cases based on changes of use case based requirement specifications," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, 2023. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2023.01401105>
- [16] R. Denaux, "Intuitive ontology authoring using controlled natural language," Ph.D. dissertation, School of Computing, University of Leeds, 2013.
- [17] K. Athiththan, S. Rovinsan, S. Sathveegan, N. Gunasekaran, K. S. A. W. Gunawardena, and D. Kasthurirathna, "An ontology-based approach to automate the software development process," *2018 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pp. 1–6, 2018.
- [18] D. Tsoukalas, M. Siavvas, M. Mathioudaki, and D. Kehagias, "An ontology-based approach for automatic specification, verification, and validation of software security requirements: Preliminary results," in *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 2021, pp. 83–91.
- [19] V. Castañeda, L. Ballejos, M. Caliusco, and M. Galli, "The use of ontologies in requirements engineering," *Journal of Researches in Engineering*, vol. 10, pp. 2–8, 01 2010.
- [20] A. A. Alsanad, A. Chikh, and A. Mirza, "A domain ontology for software requirements change management in global software development environment," *IEEE Access*, vol. 7, pp. 49 352–49 361, 2019.
- [21] H. Alrumaih, A. Mirza, and H. Alsalamah, "Domain ontology for requirements classification in requirements engineering context," *IEEE Access*, vol. 8, pp. 89 899–89 908, 2020.
- [22] L. Yang, K. Cormican, and M. Yu, "Ontology learning for systems engineering body of knowledge," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1039–1047, 2021.
- [23] E. Souza, R. Falbo, and N. Vijaykumar, "Using ontology patterns for building a reference software testing ontology," in *2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops*, 2013, pp. 21–30.
- [24] E. F. Barbosa, E. Y. Nakagawa, and J. C. Maldonado, "Towards the establishment of an ontology of software testing," in *International Conference on Software Engineering and Knowledge Engineering*, 2006.
- [25] P. Chen and A. Xi, "Research on industrial software testing knowledge database based on ontology," in *2019 6th International Conference on Dependable Systems and Their Applications (DSA)*, 2020, pp. 425–429.
- [26] L. Olsina, G. Tebes, D. Peppino, and P. Becker, "Approaches used to verify and validate a software testing ontology as an artifact," in *2020 IEEE Congreso Biental de Argentina (ARGENCON)*, 2020, pp. 1–8.
- [27] S. Banerjee, N. C. Debnath, and A. Sarkar, "An ontology-based approach to automated test case generation," *SN Computer Science*, vol. 2, no. 1, pp. 1–12, 2021.
- [28] Y. Wang, X. Bai, J. Li, and R. Huang, "Ontology-based test case generation for testing web services," in *Eighth International Symposium on Autonomous Decentralized Systems (ISADS'07)*, 2007, pp. 43–50.
- [29] A. W. Crapo and A. Moitra, "Using owl ontologies as a domain-specific language for capturing requirements for formal analysis and test case generation," in *2019 IEEE 13th International Conference on Semantic Computing (ICSC)*, 2019, pp. 361–366.
- [30] S. Ul Haq and U. Qamar, "Ontology based test case generation for black box testing," in *Proceedings of the 2019 8th International Conference on Educational and Information Technology*, ser. ICEIT 2019. New York, NY, USA: Association for Computing Machinery, 2019, p. 236–241. [Online]. Available: <https://doi.org/10.1145/3318396.3318442>
- [31] A. Jounaidi and M. Bahaj, "Designing and implementing xml schema inside owl ontology," in *2017 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2017, pp. 1–7.
- [32] O. E. Hajjamy, L. Alaoui, and M. Bahaj, "Xsd2owl2 : Automatic mapping from xml schema into owl 2 ontology," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 8, pp. 1781–1796, 2017.
- [33] N. Yahia, S. Mokhtar, and A. Ahmed, "Automatic generation of owl ontology from xml data source," *International Journal of Computer Science Issues*, vol. 9, 06 2012.
- [34] D. Pilone and N. Pitman, *UML 2.0 in a Nutshell*. O'Reilly Media, Inc., 2005.
- [35] K. M. Ting, *Precision and Recall*. Boston, MA: Springer US, 2010, pp. 781–781. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_652
- [36] J. Henarejos-Blasco, J. A. García-Díaz, O. Apolinario-Arzuabe, and R. Valencia-García, "Cnl-rdf-query: A controlled natural language interface for querying ontologies and relational databases," in *Proceedings of the 10th Euro-American Conference on Telematics and Information Systems*, ser. EATIS '20. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3401895.3402064>
- [37] R. Denaux, V. Dimitrova, A. G. Cohn, C. Dolbear, and G. Hart, "Rabbit to owl: Ontology authoring with a cnl-based tool," in *Controlled Natural Language*, N. E. Fuchs, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 246–264.

Eligible Personal Loan Applicant Selection using Federated Machine Learning Algorithm

Mehrin Anannya, Most. Shahera Khatun, Md. Biplob Hosen, Sabbir Ahmed,
Md. Farhad Hossain, M. Shamim Kaiser
Institute of Information Technology, Jahangirnagar University, Bangladesh

Abstract—Loan sanctioning develops a paramount financial dependency amongst banks and customers. Banks assess bundles of documents from individuals or business entities seeking loans depending on different loan types since only reliable candidates are chosen for the loan. This reliability materializes after assessing the previous transaction history, financial stability, and other diverse kinds of criteria to justify the reliance of the bank on an applicant. To reduce the workload of this laborious assessment, in this research, a machine learning (ML) based web application has been initiated to predict eligible candidates considering multiple criteria that banks generally use in their calculation, in short which can be briefed as loan eligibility prediction. Data from prior customers, who are authorized for loans based on a set of criteria, are used in this research. As ML techniques, Random Forest, K-Nearest Neighbour, Adaboost, Extreme Gradient Boost Classifier, and Artificial Neural Network algorithms are utilized for training and testing the dataset. A federated learning approach is employed to ensure the privacy of loan applicants. Performance analysis reveals that Random Forest classifier has provided the best output with an accuracy of 91%. Based on the mentioned prediction, the web application can decide whether the customers' requested loan should be accepted or rejected. The application was developed using NodeJs, ReactJS, Rest API, HTML, and CSS. Furthermore, parameter tuning can improve the performance of the web application in the future along with a usable user interface ensuring global accessibility for various types of users.

Keywords—*Loan eligibility prediction; machine learning; random forest; K-Nearest Neighbour; Adaboost; extreme gradient boost; artificial neural network; federated learning*

I. INTRODUCTION

People all over the world reckon on banks to gain various kinds of financial support depending on their needs. Besides, depositing individual money it provides loans to its customers assessing different conditions and criteria. In general, banks variably provide sixteen types of loan applications [1]. In recent years, the lend-leasing industry has created significant growth increasing number of individuals seeking personal loans for various purposes. This increase in demand has led to a need for more efficient and accurate methods of loan applicant selection. Loan approval criteria defer from bank to bank. Forbes refers to the top five banks in the world providing different personal loan applications and sanctioning criteria with some common attributes [2]. Assessing those top five [3][4][5][6][7] banks, it is seen that some attributes like - credit score, social security number, loan amount, loan type, mortgage information, employment, etc. are common. Depending on these criteria, traditional loan application processing carries forwards with manual reviews and human judgment which can be subjective and biased, leading to inefficient loan processing

and higher default rates consuming a huge time in taking a decision which is a cumbersome task of the banking system. Due to human error, sometimes loans are sanctioned mistakenly to some people who cannot repay banks' money with interest in proper time. Moreover, banking sectors more or less face challenges with huge data management and security issues during data processing. But the use of FL in processing all the eligible loan applicants at a time is left behind. The primary motivation behind this research is to tackle the aforementioned challenges progressively, aiming to alleviate the burden on bankers in identifying loan defaulters and streamline the loan sanction process efficiently. By providing swift decisions, this research aims to support loan applicants in making informed choices that depend on the approval of their loans. Additionally, the research aims to expedite the loan sanctioning process, reducing the waiting time for loan applicants. The research introduces a web application developed using ML and DL algorithms for selecting eligible personal loan applicants in an FL approach to ensure security and a better data management process. Since, today's modern world increasingly depends on ML for any type of big data analysis and prediction because of having different statistical models, and banks need more accurate predictive systems, in this research ML models are used for personal loan prediction. In a study [8], loan prediction has been done with a random forest algorithm providing better performance than a decision tree. Thereupon, in this research, the best accuracy-giving algorithm is selected among four ML and one DL algorithms for achieving better performance of data in checking the eligible personal loan applicants among all the submitted applications. The app uses data-driven approaches for analyzing vast amounts of data and making predictions about the candidates who are likely to be selected for the loan sanction. This leads to a more objective assessment of loan applicants and a reduced risk of loan defaults. And another lesson that has been found from analyzing different research on the loan prediction arena is, very few concrete systems have been developed for predicting eligible personal loan applicants ensuring the privacy of loan applicants. The key contributions of the research are:

- To train and test a loan prediction dataset with four ML and one DL algorithm that has been found after the literature review.
- To choose the best-performing ML algorithm among those five for loan prediction.
- To ensure the privacy, security, and robustness of data processing, an FL approach will be utilized with different loan applicant selection datasets.

- Lastly, to develop a web application for checking eligible loan applicants when customers request a loan with an online application to the bank.

The rest of the paper includes a literature review in Section II, methodology in Section III, result analysis and discussion in Section IV, and a conclusion in Section V wraps up the overall research.

II. BACKGROUND STUDY

In this section, the research background has been categorized into three subsections: ML-based loan prediction, FL-based loan prediction, and web applications using ML for loan prediction.

A. ML-based Loan Prediction

Authors in [9] used the ML approach to predict eligible candidates to receive loan amounts by collecting previous banks' data who are accredited before. For predicting loans, a simple comparative study was made in [10] based on six machine-learning classification models in R to find out whether allocating a loan to a certain person is risky or not without recommending any specific algorithm. In 2021, a comparison of seven different classifiers was performed in [11], which also showed a method for combining results from multiple classifiers. In [12], authors suggested a better method for performing the identical function in banking procedures. In terms of accuracy, it is shown in a study [13] that the Decision Tree ML algorithm outperformed rather than Logistic Regression and Random Forest ML techniques, according to the results of the trial. In [14], authors made a comparative analysis comprising Random Forest and Decision Trees, declaring the latter to have the highest accuracy when evaluated on the same dataset. To forecast an outcome on loan prediction, a Decision Tree ML algorithm was employed in [15]. In [16], Big Data mining was utilized to collect approved clients' previous data and for training and testing the ML models. Among the four ML models, the Decision tree algorithm gave the best accuracy result. In [17], three ML models are utilized to train the past data to decide whether the loan request will be accepted or not, and among them, the Decision tree algorithm outperformed than Random Forest and Logistic Regression ML approaches. An understandable artificial intelligence (AI) decision-support system was researched to automate the loan underwriting process with a belief-rule-base (BRB) and was capable of learning from and incorporating human knowledge through supervised learning, and historical data [18]. In recent times, authors of [19] made a comparative study in predicting eligible customer loan receivers using five ML algorithms recommending a Decision tree with AdaBoost ML to have the highest accuracy rate where the data cleansing mechanism played an important role. In [20], a logistic regression model was utilized for predicting the problem of forecasting loan defaulters fetching the Kaggle dataset, depending on sensitivity and specificity as the two parameters to compare the performance of the ML model. Authors of [21] used the Logistic regression model to estimate various performance metrics providing a wide range of outcomes disregarding two important variables, such as gender and marital status. A technique was utilized in [22] for developing a model using the information and outcomes of loan applicants who had already submitted applications which

discovered that the logistic regression model performs better than other models. Under the assumption that loan quality has a direct impact on a bank's profitability, in [23], a combined logistic regression method and artificial neural network (ANN) was utilized to improve the predictive performance based on real data from a rural commercial bank. In [24], a research project was made intending to create a cutting-edge algorithm to predict events for different financial institutions to protect them from fraudsters while also streamlining the pre-approval procedure for loan applications and the associated verification process. For performing data categorization with good accuracy, K-nearest neighbor (K-NN), decision tree, support vector machine, and logistic regression models are taken into account to measure their performance. A loan default dataset was used in [25], which is taken from the lending club. To address the dataset's class imbalance issue, the ADASYN (Adaptive Synthetic Sampling Approach) method was used in increasing the prediction accuracy. Following an experimental comparison, it was discovered that the fusion model proposed in this paper outperformed using three other models—Logistic Regression, Random Forest, and CatBoost—in terms of its ability to predict the likelihood of customer loan default which was trained with the dataset lowering the external risk posed by customer loan default for the online loan platform. To classify a Kaggle dataset with the best degree of feasible accuracy, it is found that the random forest classification approach provided better performance in loan candidate classification [26]. The authors of the paper [27], researched that the loan grants were given to people in previous years after mining them in their recommended model using random forest ML to predict the loan grants to develop a better risk prediction system for the network loan platform reducing its risks. In [8] also showed that Random Forest Classification outperformed better than the Decision Tree algorithm with a mean accuracy of 89.94% in finding eligible loan applicants after their loan application in a bank. Data Mining Techniques are used in [28], to assess the manual way of loan sanctions made by banks, and following that deep learning models are used to perform the task for prediction. In [29], a proprietary dataset from an agency was utilized to compare the efficacy of a variety of regression models and ML algorithms for forecasting the probability of paying the loan discovering rule-based algorithms to outperform other approaches. A model is created by Debnath et al. in [30], to forecast whether to approve credit for or deny credit utilization for clients using loan application data from consumers. The proposed model took into account the factors that affect a person's loan status and produces precise results for approving or rejecting the customer's request for credit after carefully assessing all available possibilities. To entrench the convolutional neural network (CNN) and the integration model of stacking, a loan risk prediction model called Stacking+CNN was proposed by Li et al. [31]. The prediction model created in this work was superior to the single model and other integrated models in terms of forecasting accuracy and recall rate, according to empirical results. A mechanism for foretelling loan failure was developed by Muslim et al. [32]. For the prediction analysis procedure, an enhanced light gradient boosting machine via features selection using swarm methods such as ant colony optimization and bee colony optimization was applied having a 95% success rate. Authors in [33], utilized an ML method to anticipate loan defaults recommending the Naive Bayes model to perform better than

other models. Arutjothi et al. in [34] build a credit rating model using loan status. Credit rating models are used to distinguish defaulters and legitimate consumers. This research used credit data to develop a rating model and presented an ML-based data analysis methodology with K-NN and Min-Max normalization. The proposed approach was 75.8% accurate.

B. FL-based Loan Prediction

In a recent study by [35], Yang et al., an overall description of FL with its use in different sectors like health and communication had been made. Then it found its drawback in security issues and finally discussed its future and its use in the application layer. In [36], an FL approach was utilized by Gu et al. in processing the trained model data and updating the parameters on the centralized server ensuring accuracy, privacy, and model fairness. FL approach had been remarked on by Kawa et al. in [37] for assessing credit risks by learning shared prediction models from different banks collaboratively to update their data in the central repository. Authors in [38] proposed an FL model to predict the loan requester's financial situation using the clients' banking information concluding that the F1-score metric gave identical results in both the centralized and decentralized environment. In [39], federated learning (FL) is used in finding the loan applications that have less possibility to repay the loans in due time, and a Synthetic Minority oversampling Technique (SMOTE) is used in solving the imbalanced data.

C. Web Application using ML for Loan Prediction

Sujatha et al. in [40] referred to the deployment of a web application project that utilizes an ML algorithm named logistic regression for loan prediction with a high accuracy rate. In another study by Thomas et al. [41], a similar type of suggestion has been given to achieve eligible loan applicants. But, comparisons have been made among XGBoost, K-NN, and support vector machine, recommending XGBoost to have the highest accuracy rate of about 91.6%. In a study by Shukla et al. [42], research on loan prediction-based web applications using logistic regression, random forest classification, and XGB ML algorithms has been made using Stream the lit library. The application shows either "Loan denied" or "Loan approved" status to the loan applicant customer after prediction using ML algorithms. The app can be modified to increase its accuracy in the future.

Along with the above three categories, the paper of Divate et al. [43], also predicted the outcome by mining the data of previously accepted clients. The system was developed using an AI model that delivered the most accurate result in this research. Authors in [44], employed LightGBM in predicting categorization outcomes using observational datasets as the most successful algorithm after multi-observation and multi-dimensional data cleaning. In [45], Blaszczyszynski et al. used an upgraded dataset for pre-programmed loan applications to test a tool for financial fraud prediction named DRSA-BRE and found that it performed better than existing methods. Robisco et al. Authors of [46] presented a new framework to compare ML approaches and model risk adjustments. To solve this issue, they first identified up to 13 risk variables using internal ratings-based methods, then grouped them into three primary categories: statistics, technology, and market conduct. Using

natural language processing and risk terminology based on expert knowledge, they calculated the weight of each type based on the frequency of its mentions.

The above discussion on background works assisted that myriad works prevail in the selection of eligible loan applicants using ML algorithms with good prediction providing impressive accuracy rate. But still, most of the paper indicates to increase in this accuracy rate. Moreover, web applications based on loan applicants' prediction process couldn't reach huge popularity in research sectors ensuring data security. Therefore, to develop a comprehensive web application that utilizes ML algorithms in predicting eligible loan applicants in an FL environment, further research is needed to address these challenges and ensure the fairness and transparency of the system.

III. METHODOLOGY

A. Overview

In this article, an end-to-end solution for loan prediction using ML algorithms with a series of features related to scalability, and security with a distributed federated transfer learning model has been proposed. To ensure client-side rendering with data protection, the aim is to provide a microsystem structure with exchangeable FL capabilities and client-side rendering. Elaborately, the research is working combining three parts namely - ML prediction using loan data, FL for client-side rendering, and Web application development for sanctioning loans.

B. Working Procedure

1) Web application development for sanctioning loan:

This is the main software system with whom the bankers (Admin or Bank Employees) will interact. It will work with all online loan applications from customers. The workflow of the proposed system is shown in Fig. 1. Here, a web app is developed commencing with individual access to the system. There are three types of users, namely- Admin, Customer, and Employee. The user Authentication Section will give the required roles according to the logged-in user. If the user type is Admin, then it will be redirected to the "Controls and Operates the whole system". If the user type is Customer, then the individual customers can request a loan from the bank. the system provides the loan sanctioning form to the customer. Customers fill up the form and submit it to the system. Then, the customer has to wait for its approval or rejection. If the user type is Employee, then it can view all the loan requests of the customers. When the bank employee hits the Submit button, the ML prediction analysis starts working with all the loan requests to sort the eligible loan applicants using the best ML algorithm. The process to find the best ML algorithm is shown in Fig. 2. Based on this ML prediction result, the Employee can view the customers who are accepted and rejected for the loan request. The following tools and techniques are used for its development:

- NodeJs is used for backend coding and calling the REST API using a GitHub link.
- Tensorflow javascript library is used to load and run those data.

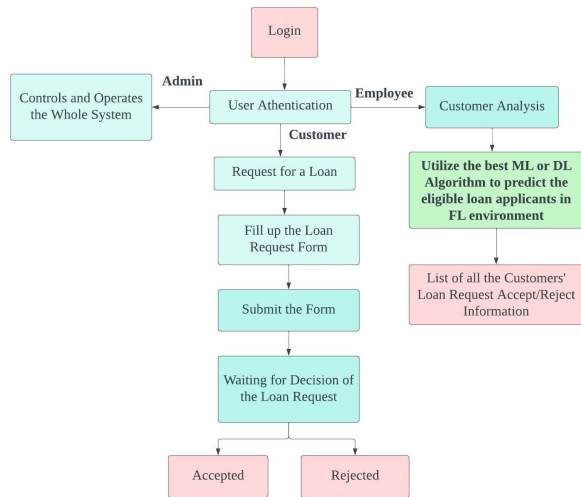


Fig. 1. Workflow diagram of the proposed system.

- After that, the generated result is uploaded to GitHub.
- Using javascript, reactjs, HTML, and CSS, the front end is developed.

2) *ML Prediction using loan data:* In the literature review section, the use of different ML algorithms has been observed, among which the ML algorithms, which are least popular, performed badly, and worked with similar datasets are chosen. Comprising all the ML and DL algorithms found to be used in similar research, four ML algorithms, namely - Random Forest Classifier, K-NN Classifier, AdaBoost Classifier, XGB Classifier, and one DL algorithm, namely - ANN have been used in this research. These five algorithms are then used to find the best one for developing the web application to predict the customers to whom the loan can be sanctioned or declined. The working procedure to find the best ML algorithm for the system is shown in Fig. 2.

At first, the data is collected from a popular dataset available on Kaggle [47]. It contains genuine 10,001 records of a bank. Then the data pre-processing has been done maintaining the following steps: *i) Null value elimination:* There are some cells in the dataset which has no values. These cells can result in improper results when tested. These null values are filled by the statistical estimation method. *ii) Label Encoding:* Some values are string-type in nature which are converted into numeric values. *iii) Correlation:* Since some attributes (LoanID, CustomerID, and Tax Liens) are not relevant to the model, this process automatically chooses useful features while removing redundant or unnecessary characteristics. Discarding a feature results in an O coefficient value. The data has 19 attributes of customers. Among these 15 attributes are used as independent attributes and 1 attribute as a dependent attribute. The attributes are given in Table I. The whole Dataset is then split into two parts: The Training Dataset and Test Dataset. All five ML models are trained with 8000 data and then tested with the rest. Then an analysis among the models has been done to select the best-performing one with the highest accuracy level. Noticeably, since ANN is a DL algorithm it is trained in the FL environment. A comparative analysis is made among ML and

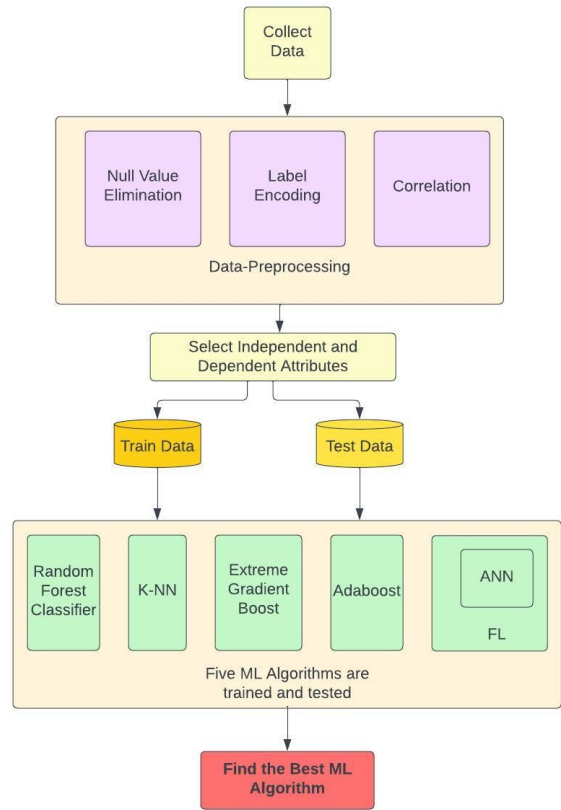


Fig. 2. Workflow diagram to find the best ML algorithm.

TABLE I. ATTRIBUTES

Dependent Attributes	Independent Attributes	
Loan Status	Current Loan Amount	Home Ownership
	Number of Open Accounts	Term
	Number of Credit Problems	Purpose
	Current Credit Balance	Credit Score
	Years of Credit History	Monthly Debt
	Maximum Open Credit	Annual Income
	Years in current job	Bankruptcies
	Months since last delinquent	

DL algorithms to choose the best one. Using the best ML or DL model, eligible customers for loan sanction are predicted and utilized in the proposed system's Utilize the best ML or DL Algorithm to predict the eligible loan applicants in the FL environment as mentioned in Fig. 1.

3) *FL for client-side rendering:* A federated learning approach is adopted to train the loan property detection model. This approach involves multiple clients, each possessing its local dataset. During each training round, the clients independently perform local training using their respective datasets. This process allows the clients to learn from their data, capturing the specific characteristics and patterns of their datasets. After the local training phase, the clients generate model updates based on their trained models. These updates typically consist of either the updated model parameters or gradients, which represent the direction and magnitude of the parameter updates. The clients then transmit their model

updates to the aggregator, a central entity responsible for coordinating the federated learning process. The federated transfer learning approach of ANN consists of several clients, where n number of client nodes $N_1, N_2, N_3 \dots \in N_n$ can participate in processing, assuming each device has at least P computational power. The local batch size B and iteration C are adaptive, depending on the user end of the generated data. Each N_i can train on private data PD and a central server-based classification model is shared between all of the nodes with a synchronized upload and download time T_u and T_d respectively. The objective function is to minimize binary cross entropy loss (BCE) over all models referring to equation 1, where y is the ground truth and ε is the prediction with optimization of all of the weight and biases denoted by w_i and b_i respectively as mentioned by Zhang et al. in [48]. This BCE loss is calculated individually for each ANN model on the client side.

$$BCE = -(y \log(\varepsilon) + (1 - y) \log(1 - \varepsilon)) \quad (1)$$

Upon receiving the model updates (BCE) from all participating clients, the aggregator computes their mean, where $\partial BCE(X_t)$ of equation 2 denotes the gradient descent and η_t is the learning rate of each node, N_i which also denotes the local update of i-th nodes with a learning convergence assumption. Again, X_t refers to the current instance of input in the i-th nodes, which can be calculated from previous instances. A large number of local models are aggregated (e.g., averaged) on the client side to create the global model. As local models are developed utilizing client-specific training data on devices, local and global models often differ. This aggregation step consolidates the model updates into a single global update, representing the collective knowledge from all the clients. By computing the mean of the model updates, the aggregator ensures a fair combination of local knowledge while preventing the dominance of any particular client. By averaging the model updates, the aggregation process balances the contributions of individual clients and facilitates the convergence towards an accurate and generalized eligible loan applicant's prediction model.

$$X_t = X_{t-1} - \sum_{i=1}^N \eta_t \partial BCE_i(X_t) \quad (2)$$

On deployment, the ML model is trained on the user side and only the prediction and updated model are sent over the network. Thus, any practical or private information needed for the ML model to operate will be separated from the central cloud storage, resulting in a more secure and reliable application system. It also solves critical issues like data security, privacy, and authorized access. This is also a more decentralized approach where edge devices actively participate in computation, reducing the computation complexity on a central server. Therefore, this process also enables reducing the unnecessary model parametric complexity.

IV. RESULT ANALYSIS AND DISCUSSION

This research can determine the eligibility of a customer to get a loan. After getting all the related information about customers, the system checks that using the best ML algorithm, the system can approve or reject the loan applicants.

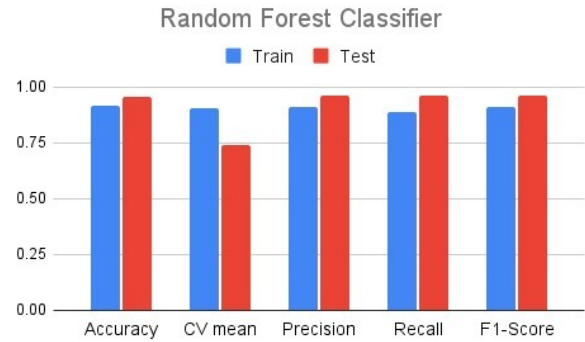


Fig. 3. Performance comparison of random forest classifier.

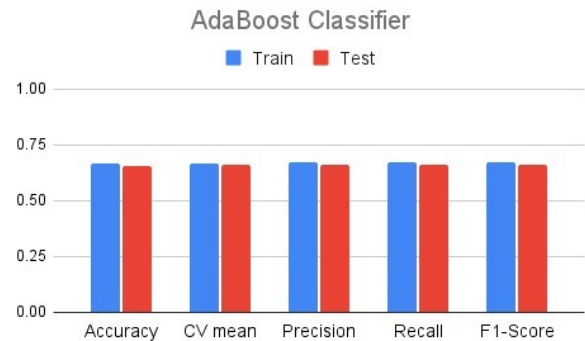


Fig. 4. Performance comparison of Adaboost classifier.

A. ML Prediction Result

To determine the best-performing ML algorithm for the dataset [47], accuracy, CV Mean, Precision, Recall, and F1-score metrics of the confusion matrix are used since most of the research papers referred to in the literature review section used them. The graphs for each of the five algorithms using the above-mentioned metrics are analyzed here:

1) *Random forest classifier for loan prediction:* Fig. 3 describes that the test data performed better than the training data generating a value near 1 for each of the attributes except for the CV mean. CV means calculated a significant degradation in value compared to all the train and test data. All the metrics of the train data are generating a value near 0.9 except for recall which is slightly below compared to the other metrics.

2) *Adaboost classifier for loan prediction:* From Fig. 4, it is observed that the training data performed better than the test data for all metrics and are near 0.65 which is poor than the Random Forest Classifier in Fig. 3.

3) *K-NN classifier for loan prediction:* Fig. 5, Fig. 6, Fig. 7, and Fig. 8 describe graphs for four different values of k. Here, K=7, 11, 13 and, 17 were used to of K-NN identify any significant change in its pattern. When the value of K in Fig. 5 was 7, it was observed that for all the metrics of confusion matrix, the probability was above 0.75 except for CV-mean. But, when the value of K in Fig. 6 was bit increased to 11, it was observed that for all the metrics of confusion matrix,

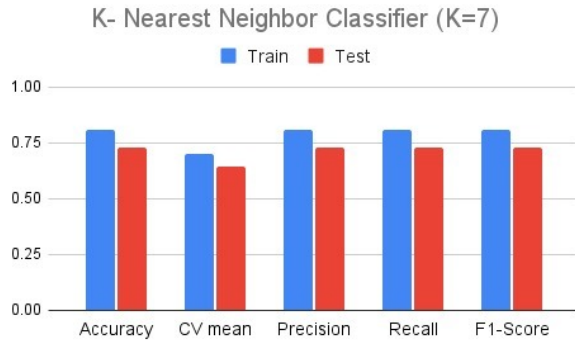


Fig. 5. Performance comparison of K-NN classifier (n=7).

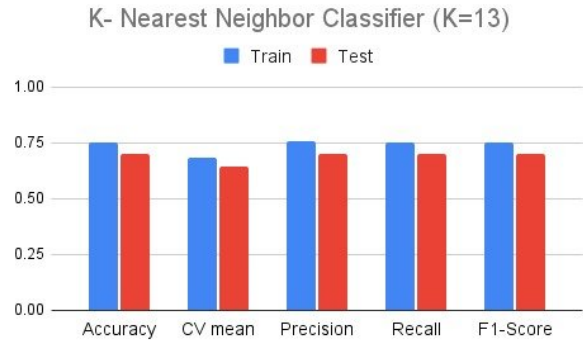


Fig. 7. Performance comparison of K-NN classifier (n=13).

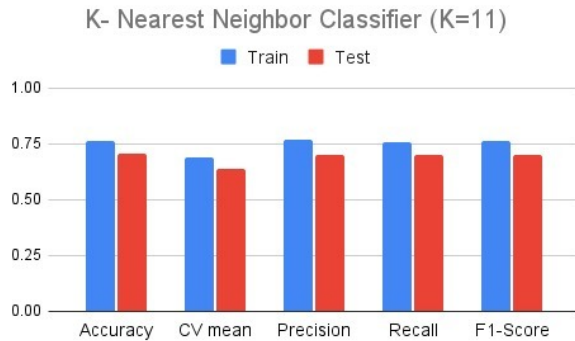


Fig. 6. Performance comparison of K-NN classifier (n=11).

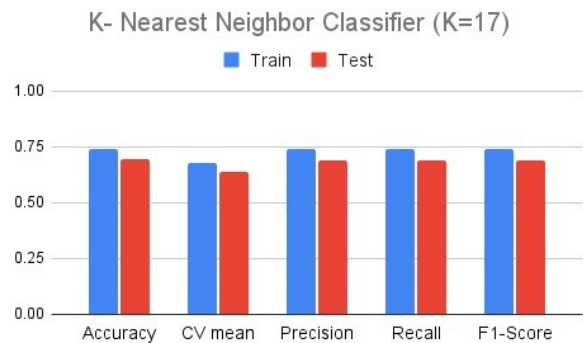


Fig. 8. Performance comparison of K-NN classifier (n=17).

the probability was 0.75 except for CV-mean. Similarly, when the value of K in Fig. 7 was a bit increased from 11 to 13, it was observed that for all the metrics of confusion matrix, the probability was 0.75 except for CV-mean. However, when the value of K in Fig. 8 was searched covering wide range to 17, it was observed that for all the metrics of confusion matrix, the probability was near to 0.75 except for CV-mean. Furthermore, no significant differences were observed in it. For all the four values of k, the training data performed better than test data similar to Adaboost Classifier in Fig. 4 calculating a value around 0.75 for each of the metrics with a significant decrease in the value of CV Mean metrics. But the values are less than the Random Forest Classifier in Fig. 3.

4) *Extreme gradient boosting classifier*: Fig. 9 describes that the training data performed slightly better than the test data calculating a value near 0.60 for each of the metrics similar to the Adaboost Classifier in 4 and K-NN in Fig. 5 to 8. But couldn't outrange the Random Forest Classifier in Fig. 3.

5) *ANN*: This section generates Fig. 10 and Fig. 11 using the equation 2 and 1 respectively. Since it uses a neural network to perform the calculation, with the increase in the number of epochs [49] i.e. the learning rate, observing the Fig. 10 and Fig. 11, it is seen that accuracy of FL-based ANN is also increasing for both the training and the testing data with a corresponding decrease in loss value. But since the measurement is made on a scale of 1, the peak value of it is around 0.8 which is less than the Random Forest Classifier in Fig. 3. Considering all the values of each confusion matrix for

all the ML and DL algorithms, a comparative graph is created in Fig. 12. In this graph, for the overall analysis, only the accuracy and F1-score metrics are selected for both training and test data since they gave excellent results for all the ML algorithms. However, the FL-based ANN couldn't beat the ML algorithm even after having multiple iterations. Hence, it is concluded that Random Forest Classifier's performance is the best in comparing all the other ML and DL algorithms. This Random Forest Classifier is then used in the FL environment for data analysis of the web application.

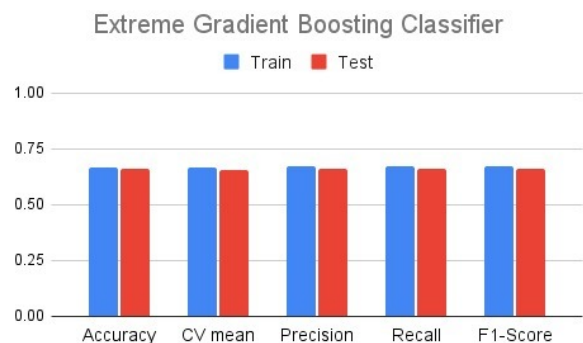


Fig. 9. Performance comparison of XGB classifier model

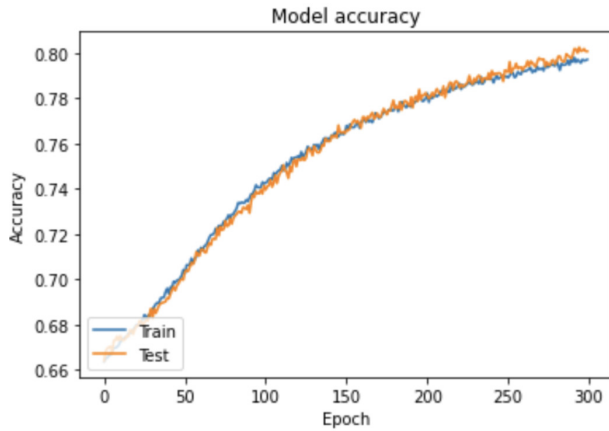


Fig. 10. ANN classifier model accuracy.

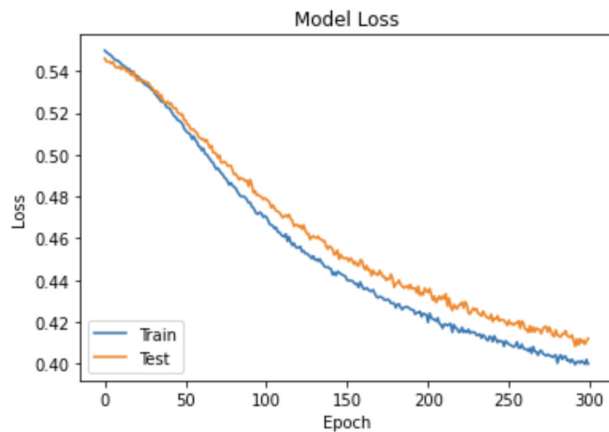


Fig. 11. ANN classifier model loss.

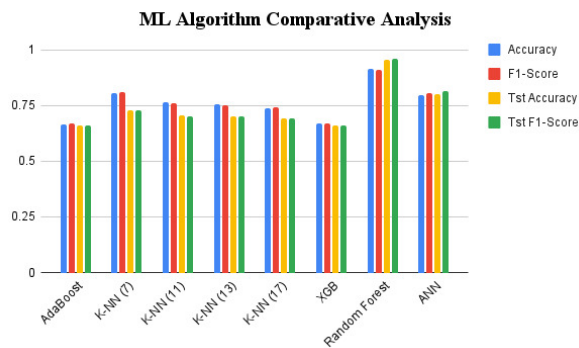


Fig. 12. Comparative analysis of ML algorithms.

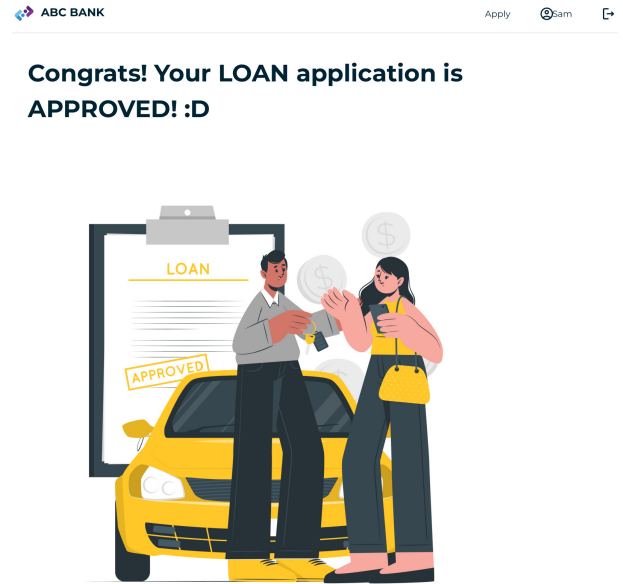


Fig. 13. Loan approval UI of a customer.

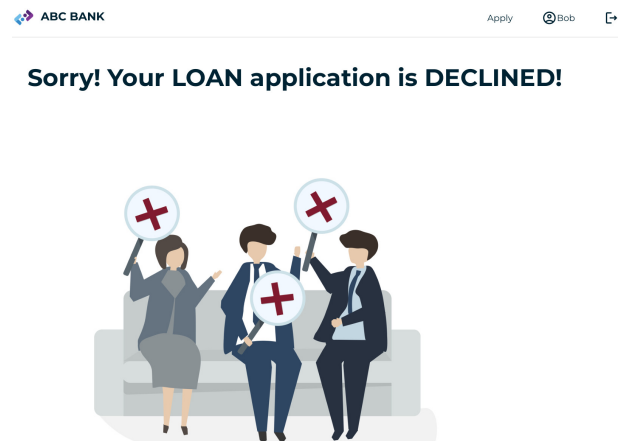


Fig. 14. Loan decline UI of a customer.

B. User Interface (UI) of Web Application

In this section, some salient figures of the developed web application have been highlighted using which the customer and bank employee will interact for loan processing. Here, ABC Bank is considered an exemplary name of a bank.

- **User-Customer:** Fig. 13 shows a customer named Sam has been sanctioned with his requested loan and Fig. 14 shows a loan decline UI for a customer named Bob. However, the customer's application form's UI is skipped from inclusion.
- **User-Employee:** Fig. 15 shows the employee dashboard UI which comes after processing the customers' loan application using ML prediction techniques. In Fig. 15, it is seen that the customer with LoanID: 1 is declined from getting the loan, LoanID: 2 has been approved for loan sanction, and LoanID: 3 and 4's loan requests are still on review status. Employees can review loan requests using ML algorithms. If the

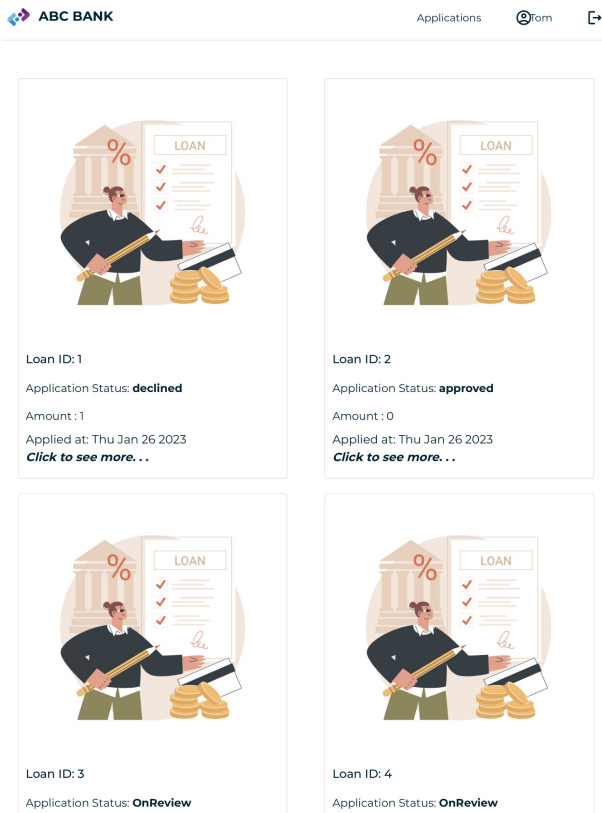


Fig. 15. Employee dashboard.

APPLICATION INFORMATION	
Attributes	Value
Name	Sam
Customer ID	5
Email	sam@gmail.com
Bank Account No.	54312345
Loan Amount	0
Annual Income	70000
Bankruptcies	0
Credit Balance	100001
Credit Score	1200
Last Delinquent	1000
Max Open Credit	1000000
Monthly Debt	0
Open Accounts	1
Purpose	3
Term	1
Years in Current Job	6
Years of Credit History	50
Account Created	Thu Jan 26 2023
Application date	Thu Jan 26 2023
Prediction Score	94 %
Application Status	approved

Fig. 16. Employee view of the loan approved customer.

employee wants to view the detailed information of the loan applicants then the UI regarding that is shown in Fig. 16 gives an approved loan applicant's details and Fig. 17 gives a declined loan applicant's details. Elaborately, both the figures show all the information of the customer to whom the loan has been sanctioned and whose loan application is rejected respectively using ML algorithms.

APPLICATION INFORMATION	
Attributes	Value
Name	Bob
Customer ID	4
Email	bob@gmail.com
Bank Account No.	2134567890
Loan Amount	1
Annual Income	0
Bankruptcies	0
Credit Balance	1
Credit Score	1
Last Delinquent	1
Max Open Credit	1
Monthly Debt	1
Open Accounts	1
Purpose	1
Term	1
Years in Current Job	1
Years of Credit History	0
Account Created	Thu Jan 26 2023
Application date	Thu Jan 26 2023
Prediction Score	0 %
Application Status	declined

Fig. 17. Employee view of the loan declined customer.

V. CONCLUSION

Sanctioning a loan is a challenging task for bankers since there occur some phenomena when the borrowers cannot return their debts in due time. Sometimes, debt cannot be collected too due to some misjudgment. Various types of loans are provided by the banks. In this research, an ML-based web application has been used to check the eligibility of personal loan applicants. To conduct the task, data is used for prediction using four ML and one DL algorithm. The prediction has been performed depending on some attributes in which the most crucial factors that are considered in taking decisions are - loan amount, loan length, loan term, and age. Among those five ML algorithms, Random Forest Classifier has been suggested to be used by the banks since it has given the best result for all the metrics of the confusion matrix. Moreover, another remarkable component of the research is the implementation of a decentralization technique in local PC for data processing using the FL approach to ensure its data security and robustness.

However, the research lacks working with more real and relevant data that can effect the accuracy augmentation of the ml algorithms. It could have worked with more latest ml algorithms which have not been used in this type of research. The back-end architecture of the web application have been developed with modern programming tools.

In future, the research could have work with more real data integrating more empirical attributes that the banks follow and use during their assessment so that the accuracy of prediction can be enhanced. Furthermore, the user interface of the web application can also be enhanced in the future using modern tools and techniques.

REFERENCES

- [1] Forbes, *16 Types of Loans to Help You Make Necessary Purchases*, <https://www.forbes.com/advisor/loans/types-of-loans/>, Accessed: 18 January 2023.
- [2] Forbes, *Top 5 International Banks for Worldwide Banking*, <https://statrys.com/blog/best-international-banks/>, Accessed: 18 January 2023.

- [3] Citi, *APersonal Lines & Loans - See all lines & loans from Citibank® - Citibank*, <https://online.citi.com/US/ag/personal-loan?intc=17505LOBAPersonalLoansPos4/>, Accessed: 18 January 2023.
- [4] HSBC, *Personal Loan — HSBC Bank Bangladesh*, <https://www.hsbc.com.bd/1/2/retail-banking/loans/personal-loan/>, Accessed: 18 January 2023.
- [5] Capital One, *Let's get personal: Understanding how to get a personal loan*, <https://www.capitalone.com/bank/money-management/banking-basics/how-to-get-a-personal-loan/>, Accessed: 18 January 2023.
- [6] JPMorgan Chase & Co., *AChoosing the Right Mortgage Loan — Chase*, <https://www.chase.com/personal/mortgage/mortgage-purchase/choose-a-loan/>, Accessed: 18 January 2023.
- [7] Thomas (TJ) Porter, *How to Get Bank of America Personal Loans 2022*, <https://www.mybanktracker.com/personal-loans/faq/how-to-get-bank-of-america-personal-loans-268385/>, Accessed: 18 January 2023.
- [8] Kumar, Narra Rahul, and L. Rama Parvathy, *Higher Accuracy on Loan Eligibility Prediction using Random Forest Algorithm over Decision Tree Algorithm*, *Baltic Journal of Law & Politics* 15, no. 4 (2022): 241-251.
- [9] Gupta, Anshika, Vinay Pant, Sudhanshu Kumar, and Praveesh Kumar Bansal, *Bank Loan Prediction System using Machine Learning*, In 2020 9th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 423-426. IEEE, 2020.
- [10] Arun, K., Ishan, G. and Sanmeet, K., *Loan approval prediction based on machine learning approach*, 2016. *IOSR J. Comput. Eng.*, 18(3), pp.18-21.
- [11] Luczak, Aleksandra, Maria Ganzha, and Marcin Paprzycki, *Probability of Loan Default—Applying Data Analytics to Financial Credit Risk Prediction*, In *Intelligent Systems, Technologies and Applications: Proceedings of Sixth ISTA 2020, India*, pp. 1-16. Springer Singapore, 2021.
- [12] Ghildiyal, Bhawana, Shubham Garg, and Vivek Raturi, *Analyze of different algorithms of machine learning for loan approval*, In *Smart Trends in Computing and Communications: Proceedings of SmartCom 2021*, pp. 719-727. Springer Singapore, 2022.
- [13] Tejaswini, J., T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni, and Venkata Rao Maddumala, *ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH*, *Journal of Engineering Science* 11, no. 4 (2020): 523-532.
- [14] Madaan, Mehul, Aniket Kumar, Chirag Keshri, Rachna Jain, and Preeti Nagrath, *Loan default prediction using decision trees and random forest: A comparative study*, In *IOP Conference Series: Materials Science and Engineering*, vol. 1022, no. 1, p. 012042. IOP Publishing, 2021.
- [15] Gomathy, C. K., Ms Charulatha, Mr Aakash, and Ms Sowjanya, *The Loan Prediction Using Machine Learning.*, *International Research Journal of Engineering and Technology* 8, no. 10 (2021).
- [16] Supriya, Pidikiti, Myneedi Pavani, Nagarapu Saisushma, Namburi Vimala Kumari, and K. Vikas, *Loan prediction by using machine learning models*, *International Journal of Engineering and Techniques* 5, no. 2 (2019): 144-147.
- [17] Singh, Vishal, Ayushman Yadav, Rajat Awasthi, and Guide N. Partheeban, *Prediction of modernized loan approval system based on machine learning approach*, In 2021 International Conference on Intelligent Technologies (CONIT), pp. 1-4. IEEE, 2021.
- [18] Sachan, Swati, Jian-Bo Yang, Dong-Ling Xu, David Eraso Benavides, and Yang Li, *An explainable AI decision-support-system to automate loan underwriting*, *Expert Systems with Applications* 144 (2020): 113100.
- [19] Kumar, Ch Naveen, D. Keerthana, M. Kavitha, and M. Kalyani, *Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector*, In 2022 7th International Conference on Communication and Electronics Systems (ICCES), pp. 1007-1012. IEEE, 2022.
- [20] Sheikh, Mohammad Ahmad, Amit Kumar Goel, and Tapas Kumar, *An Approach for Prediction of Loan Approval using Machine Learning Algorithm*, In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 490-494. IEEE, 2020. Kath2021
- [21] Pramod, Ms Kathe Rutika, Ms Dapse Punam Laxman, Ms Panhale Sakshi Dattatray, and Ms Avhad Pooja Prakash, *Prediction of Loan Approval using Machine Learning Algorithm: A Review Paper*, (2021).
- [22] Rath, Golak Bihari, Debasish Das, and BiswaRanjan Acharya, *Modern approach for loan sanctioning in banks using machine learning*, In *Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019*, pp. 179-188. Springer Singapore, 2021.
- [23] Li, Yiheng, and Weidong Chen, *Entropy method of constructing a combined model for improving loan default prediction: A case study in China*, *Journal of the Operational Research Society* 72, no. 5 (2021): 1099-1109.
- [24] Hemachandran, Kannan, Raul V. Rodriguez, Rajat Toshniwal, Mohammed Junaid, and Laxmi Shaw, *Performance analysis of different classification algorithms for bank loan sectors*, In *Intelligent Sustainable Systems: Proceedings of ICISS 2021*, pp. 191-202. Singapore: Springer Singapore, 2021.
- [25] Li, Xingyun, Daji Ergu, Di Zhang, Dafeng Qiu, Ying Cai, and Bo Ma, *Prediction of loan default based on multi-model fusion*, *Procedia Computer Science* 199 (2022): 757-764.
- [26] Bhanu, L., and Dr S. Narayana, *Customer Loan Prediction Using Supervised Learning Technique*, *International Journal of Scientific and Research Publications* 11, no. 6 (2021): 78.
- [27] Pandimurugan, V., D. Usha, M. Nageswara Guptha, and M. S. Hema, *Random forest tree classification algorithm for predicating loan*, *Materials Today: Proceedings* 57 (2022): 2216-2222.
- [28] Pradeep Kumar Singh , Zdzislaw Polkowski Sudeep Tanwar, Sunil Kumar Pandey Gheorghe Matei, Daniela Pirvu, *Innovations in Information and Communication Technologies*, Proceedings of International Conference on ICRiHE - 2020, Delhi, India: IICT-2020.
- [29] Bellotti, Anthony, Damiano Brigo, Paolo Gambetti, and Frédéric Vrms, *Forecasting recovery rates on non-performing loans with machine learning*, *International Journal of Forecasting* 37, no. 1 (2021): 428-444.
- [30] Debnath Bhattacharyya, N. Thirupathi Rao, *Machine Intelligence and Soft Computing*, Proceedings of ICMISC 2020.
- [31] Li, Meixuan, Chun Yan, and Wei Liu, *The network loan risk prediction model based on Convolutional neural network and Stacking fusion model*, *Applied Soft Computing* 113 (2021): 107961.
- [32] Muslim, Much Aziz, Yosza Dasril, Muhammad Sam'an, and Yahya Nur Ifriza, *An improved light gradient boosting machine algorithm based on swarm algorithms for predicting loan default of peer-to-peer lending*, *Indonesian Journal of Electrical Engineering and Computer Science* 28, no. 2 (2022): 1002-1011.
- [33] Kadam, Ashwini S., Shraddha R. Nikam, Ankita A. Aher, Gayatri V. Shelke, and Amar S. Chandgude, *Prediction for loan approval using machine learning algorithm*, *International Research Journal of Engineering and Technology (IRJET)* 8, no. 04 (2021).
- [34] Arutjothi, G., and C. Senthamarai, *Prediction of loan status in commercial bank using machine learning classifier*, In 2017 International Conference on Intelligent Sustainable Systems (ICISS), pp. 416-419. IEEE, 2017.
- [35] Yang, Aimin, Zezhong Ma, Chunying Zhang, Yang Han, Zhibin Hu, Wei Zhang, Xiangdong Huang, and Yafeng Wu., *Review on application progress of federated learning model and security hazard protection*, *Digital Communications and Networks* 9, no. 1 (2023): 146-158.
- [36] Gu, Xiuting, Zhu Tianqing, Jie Li, Tao Zhang, Wei Ren, and Kim-Kwang Raymond Choo., *Privacy, accuracy, and model fairness trade-offs in federated learning*, *Computers & Security* 122 (2022): 102907.
- [37] Kawa, Deep, Sunaina Punyani, Priya Nayak, Arpita Karkera, and Varshapriya Jyotinagar., *Credit risk assessment from combined bank records using federated learning*, *International Research Journal of Engineering and Technology (IRJET)* 6, no. 4 (2019): 1355-1358.
- [38] Imteaj, Ahmed, and M. Hadi Amini, *Leveraging asynchronous federated learning to predict customers financial distress*, *Intelligent Systems with Applications* 14 (2022): 200064.
- [39] Shingi, Geet, *A federated learning based approach for loan defaults prediction*, In 2020 International Conference on Data Mining Workshops (ICDMW), pp. 362-368. IEEE, 2020.
- [40] Sujatha, C. N., Abhishek Gudipalli, Bh Pushyami, N. Karthik, and B. N. Sanjana., *Loan Prediction Using Machine Learning and Its Deployment On Web Application*, In 2021 Innovations in Power and Advanced Computing Technologies (i-PACT), pp. 1-7. IEEE, 2021.
- [41] Thomas, T. C., J. P. Sridhar, M. J. Chandrashekar, Makarand Upadhyaya, and Sagaya Aurelia., *Developing a website for a bank's Machine Learning-Based Loan Prediction System*.

- [42] Shukla, Saurabh, Arushi Maheshwari, and Prashant Johri., *Comparative analysis of ml algorithms & stream lit web application*, In 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), pp. 175-180. IEEE, 2021.
- [43] Diwate, Yash, Prashant Rana, and Pratik Chavan, *Loan Approval Prediction Using Machine Learning*, International Research Journal of Engineering and Technology (IRJET) 8, no. 05 (2021).
- [44] Ma, Xiaojun, Jinglan Sha, Dehua Wang, Yuanbo Yu, Qian Yang, and Xueqi Niu, *Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning*, Electronic Commerce Research and Applications 31 (2018): 24-39.
- [45] Błaszczyński, Jerzy, Adiel T. de Almeida Filho, Anna Matuszyk, Marcin Szelag, and Roman Słowiński, *Auto loan fraud detection using dominance-based rough set approach versus machine learning methods*, Expert Systems with Applications 163 (2021): 113740.
- [46] Alonso Robisco, Andrés, and Jose Manuel Carbo Martinez, *Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction*, Financial Innovation 8, no. 1 (2022): 70.
- [47] Kaggle, *Classificação - DataSet*, <https://kaggle.com/code/jcaminha/classificacao-dataset/>, Accessed: 18 January 2023.
- [48] Zhang, Zhilu, and Mert Sabuncu., *Generalized cross entropy loss for training deep neural networks with noisy labels*, Advances in neural information processing systems 31 (2018).
- [49] Deepchecks, *Epoch in Machine Learning*, <https://deepchecks.com/glossary/epoch-in-machine-learning/>, Accessed: 18 January 2023.

A Low-Cost Wireless Sensor System for Power Quality Management in Single-Phase Domestic Networks

Cristian A. Aldana B, Edison F. Montenegro A
Universidad Distrital, Francisco José de Caldas, Bogotá D.C., Colombia

Abstract—This article presents a novel low-cost hardware and software tool for monitoring power quality in single-phase domestic networks using an ESP32 microcontroller. The proposed embedded system allows remote evaluation and monitoring of electrical energy consumption behavior through non-invasive current measurement parameters. Based on these measurements, power, power factor, total harmonic distortion, and energy consumption are calculated. The collected data is then published and visualized on a free and open IoT application in the cloud. The tool was designed to be both cost-effective and high-quality. During laboratory testing, the equipment demonstrated a high level of precision, as compared to a network analyzer. Additionally, the design utilized the smallest number of components possible, while still maintaining quality performance. The ESP32 microcontroller enables wireless data transmission, making remote monitoring and management of energy consumption more accessible and efficient. Moreover, the non-invasive measurement method makes the tool safer and more user-friendly, as it does not require any interruption of power supply. The proposed tool can help identify and address power quality issues that arise in domestic networks, which can have a significant impact on energy consumption and costs. The IoT application enables users to access their power consumption data remotely, facilitating better energy management and reducing wastage.

Keywords—Cost-effective; current measurement; energy consumption; ESP32 microcontroller; non-invasive; power quality; remote monitoring

I. INTRODUCTION

As the world population continues to grow, so does the demand for energy. Unfortunately, this increase in demand is accompanied by a decline in natural resources and an increase in environmental pollution, leading to climate change. It is crucial to raise awareness among people about the importance of energy usage and to bridge the gaps in accessibility and culture [1], to create a better environment for all.

However, the current home energy measurement technology falls short of achieving this goal. The technology is primarily used to charge users for the service provided, without giving them the means to measure, verify, or control the amount charged. Furthermore, most users do not understand or have access to information provided by their energy meter [2], as it typically displays only a set of numbers on a counter located outside their homes.

Modern and innovative technologies such as embedded systems with high-capacity microcontrollers and IoT information technologies can provide more efficient, compact, and cost-effective solutions for domestic energy measurement [3].

These technologies have the potential to optimize energy measurement, enabling more precise remote monitoring and control from anywhere in the world, and providing the user with easily accessible and understandable information [4]. By creating a more precise and accessible energy measurement tool, users can plan their energy consumption more intelligently and responsibly. This can help mitigate the effects of global warming and promote energy efficiency [5], thus contributing to the conservation of the environment.

The aim of this research is to design and implement a robust IoT meter prototype capable of measuring electrical energy consumption and providing the average user with accurate real-time information about their energy usage at home. This information will enable the user to manage their energy consumption with an innovative and cost-effective solution, contributing to energy efficiency and promoting sustainability [6].

The proposed tool provides a high-quality and efficient solution for domestic networks, enabling users to be more aware of their energy consumption and manage it more intelligently and responsibly. By creating a robust and user-friendly IoT meter, this research aims to address the issues of accessibility and awareness in the use of energy. The use of innovative technologies such as embedded systems and microcontrollers has the potential to revolutionize the field of power quality monitoring and management [7], contributing to the sustainable development of modern society.

A. Issues in Existing Work

In recent years, power quality management in single-phase domestic networks has witnessed significant advancements. However, there remain certain inherent challenges. The majority of existing systems fall into one of the following categories:

- 1) Systems that prioritize advanced functionalities but, as a result, become too complex and expensive for regular household users [8], [9].
- 2) Systems that are affordable but compromise on the depth of data and the quality of measurements they provide. This often leads to a lack of complete understanding of energy usage patterns, reducing the efficacy of management efforts [10].
- 3) Devices that focus solely on energy consumption, overlooking the broader aspect of power quality which is crucial given the increased presence of non-linear loads in contemporary households [11].

- 4) Systems that although being technologically advanced, aren't user-friendly or accessible for an average user. Such systems, despite their potential, have limited real-world application due to the steep learning curve they present [12].

These issues represent significant barriers for users who wish to adopt sustainable energy consumption practices in their homes.

In addition to its simplicity and cost-effectiveness, the prototype was designed to maintain quality in both hardware and software, while keeping in mind the experience and prices of the market. The hardware and software were designed to ensure flexibility and openness to the public. In fact, there is no need for subscription to IoT platform Adafruit.io, which allows data to be uploaded at a rate of one data per second and stored at no cost.

The ESP32 Dev Kit v1 microcontroller [13] serves as the brain of the circuit and has an integrated Wi-Fi capability, which is a significant advantage for sending data directly to the cloud without additional devices. Wi-Fi enables the data to be sent in real-time to the internet or a local network if required. In this case, it is connected to the cloud platform dashboard Adafruit.io. The SCT 013030 [14] current sensor is used for signal conditioning, which has a ratio of 30A/1Vac, is ideal for its low cost and suitability for many households that do not exceed a power consumption value of 7200 W. One of its main features is its non-invasive type, which allows the installation to be safer by avoiding the need to cut parts of the wiring. The current conditioning is carried out through a precision full-wave rectification circuit to obtain a reliable signal at the ADC input and with minimal voltage loss in the diodes. An operational amplifier, the LM324, is used for this purpose. With these materials, the prototype can be constructed. The rest of the design is software-based and is programmed in the Arduino IDE environment, which is widely known and easily programmable using the C++ language.

Upon analyzing various studies related to energy meters, it was observed that the technologies used do not meet the characteristics of simplicity in their design. This means that more than one board must be used to fulfill the same functions that can be provided by an ESP32 microcontroller, which has greater processing capabilities and is more integrated and cheaper than an Arduino solution with an additional IoT communication card. Additionally, many of the designs do not consider the price, which can be a significant barrier to the acquisition of an energy meter by the user. Although some functions, such as bidirectional measurement, may seem important, most people currently do not have access to this type of technology due to their low-income status, so it would not make sense to include this function [15]. Other technologies are based on conditioning additional circuits to the energy meter. However, if the energy meter makes a mistake in the measurement, the additional device will also be incorrect, which does not provide reliability to the readings. Many designs also include functions of an incorporated power analyzer, which seems relevant given the increase of non-linear loads in homes and which can shed light on their impact on the distribution network. However, the contemplated design is not compact, and the solution is not cheap.

B. Overcoming the Challenges with the Proposed Approach

Our research aims to bridge the aforementioned gaps in power quality management for single-phase domestic networks. The proposed wireless sensor system addresses the need for a balance between advanced functionalities and user accessibility. Leveraging the capabilities of the ESP32 microcontroller, the system offers a comprehensive suite of measurements, from active power to total harmonic distortion, while ensuring that the data is readily accessible through an intuitive IoT interface [16]. Furthermore, our commitment to a cost-effective design ensures that our solution remains affordable, promoting widespread adoption and contributing substantially to the global energy efficiency movement.

Over recent years, the significance of monitoring energy consumption has grown considerably, especially in the wake of rising energy demands and the increased focus on sustainable living. As a response to these trends, this work introduces a comprehensive IoT-based current meter, designed to provide real-time insights into energy consumption patterns, thus facilitating better energy management. Our endeavor is rooted in the following main contributions:

- Development of an IoT-based current meter firmware that harnesses the power of the Fast Fourier Transform (FFT) for precise and efficient current monitoring.
- Configuration of the Analog-to-Digital Converter (ADC) tailored to ensure a detailed representation of the current waveform, enhancing energy monitoring capabilities.
- Adoption of Robin Scheibler's FFT library for high fidelity signal decomposition.
- Empirical derivation of an amplitude correction factor, thereby refining energy consumption measurements.
- Comprehensive power and energy calculations, offering real-time energy consumption insights.
- Incorporation of Total Harmonic Distortion of Current (THDi) calculations, revealing the system's performance metrics and potential energy consumption anomalies.
- Rigorous prototype testing and validation against industry-standard measurement tools, ensuring the reliability and accuracy of the developed system.

The ensuing sections detail the methods employed, the design considerations, and the empirical findings that validate the contributions outlined above.

II. RELATED WORKS

The efficient use of electrical energy is becoming increasingly important in the face of rising demand and limited resources [17]. Energy providers charge customers for the energy delivered to their homes or businesses, but not all of this energy is utilized efficiently; a portion is wasted. The energy demand of a system is known as the *apparent power* or *absorbed power*, which can be further broken down into the *active power* that is actually used and the *reactive power* that is wasted. In practice, active power should be as close

as possible to apparent power, but this is not always the case. The difference between the two can be measured by the power factor [18]. With the growing number of electronic devices in homes, the situation is different, as loads have increased their nonlinear components, increasing the harmonic content and the power factor, which can significantly affect the network and loads [11]. Therefore, it is essential to begin measuring these variables in homes to facilitate future studies on the impact of harmonics in the distribution network. The proposed tool can help identify and address power quality issues that arise in domestic networks, which can have a significant impact on energy consumption and costs.

In recent years, several studies and devices have been developed to measure power consumption and other parameters in order to determine the quality of energy in homes. Trujillo and Lorenzo [8] proposed an electric power consumption analyzer using Arduino and MATLAB to study various household loads. Benalcazar et al. [9] analyzed the generation and correction of the main sources of harmonic distortion commonly found in household and some industrial electrical networks, using LabVIEW and the National Instruments DAQ 6008 acquisition card. Garcia-Granados et al. [19] designed and implemented a single-phase power analyzer for domestic use using voltage, current, and power waveforms.

In the same vein, Mathew [10] aimed to optimize energy consumption by implementing intelligent control of household appliances using a smart meter and IoT. The system analyzed usage patterns, collected physical variables through an Arduino Uno, and featured a switching mechanism. Furthermore, the system was parametrized based on peak demand hours to reduce the electricity bill.

The disadvantages of the prototypes compared to the one proposed are mainly related to their limitations in terms of functionality, precision, and cost-effectiveness. For instance, the prototype developed in [8] only measures the active power consumption, whereas the proposed system using an ESP32 can measure not only the active power but also other parameters such as power factor, total harmonic distortion, and energy consumption. Similarly, the prototype presented in [9] focuses on the analysis of harmonic distortion in household and industrial networks but lacks the capability to provide real-time energy consumption information to users. The system presented in [19] is a single-phase power analyzer that can measure voltage, current, and power, but it does not have wireless data transmission capabilities and requires additional hardware for data visualization. Additionally, the prototypes mentioned in the paragraph do not take advantage of modern and innovative technologies such as IoT, which can facilitate remote monitoring and management of energy consumption in a cost-effective and user-friendly way.

Several prototypes and devices have been developed to monitor and control energy consumption in households using IoT technology. Ramani et al. [12] propose a prototype that combines IoT with solar energy monitoring and household energy consumption control using an Arduino and an ESP8266. The aim of the prototype is to improve energy efficiency by controlling the use of generated and consumed energy in parallel. However, the use of an ESP8266 may limit the range of the system due to its lower wireless transmission capability compared to the ESP32.

Another prototype proposed by Sheeba et al. [20] improves the conventional digital energy meter by converting it into an IoT-enabled device using an optocoupler circuit to capture LED pulses and a system that sends data to a cloud-based platform called Firebase. This design reduces the number of components required and improves the efficiency of the existing infrastructure.

Although several commercial energy meters are available in the market, they have limitations. For example, eMon energy [21] is an energy monitoring system in Indonesia that can measure up to 14 circuits accurately, store and upload real-time information to cloud services such as InfluxDB & Grafana, Emon CMS, or Mango Automation. However, it may not be suitable for all households as it may not measure voltage in all phases or may not be configurable for both residential and industrial sectors.

The use of intelligent meters for measuring energy consumption has gained significant attention in recent years. One such example is the Engage smart meter [22], which can measure instantaneous values with power factor in four quadrants. It operates as a bidirectional energy meter for both consumption and generation, with a cost of approximately 141.99 Euros. Another example is the Wibeec Box [23], a WiFi-enabled electricity meter that monitors data on electricity consumption and allows users to view it from their smartphones, tablets, or computers. This device has a cost of 181.75 Euros. These devices provide valuable information on energy consumption and can help users make informed decisions about their energy usage, leading to energy savings and a reduction in greenhouse gas emissions. However, they often come with a high price tag, which can be a barrier to their widespread adoption.

To address this issue, we propose a novel low-cost hardware and software tool for monitoring power quality in single-phase domestic networks using an ESP32 microcontroller [16]. Our embedded system enables remote evaluation and monitoring of electrical energy consumption behavior through non-invasive current measurement parameters. Based on these measurements, power, power factor, total harmonic distortion, and energy consumption are calculated. The collected data is then published and visualized on a free and open IoT application in the cloud.

Our tool was designed to be both cost-effective and high-quality, utilizing the smallest number of components possible while still maintaining quality performance. The ESP32 microcontroller enables wireless data transmission, making remote monitoring and management of energy consumption more accessible and efficient. Moreover, the non-invasive measurement method makes the tool safer and more user-friendly, as it does not require any interruption of power supply. Our proposed tool can help identify and address power quality issues that arise in domestic networks, which can have a significant impact on energy consumption and costs. The IoT application enables users to access their power consumption data remotely, facilitating better energy management and reducing wastage.

III. METHODS

In this section, we outline the materials utilized for the hardware design, as well as the software employed in the

development and implementation of the prototype. Our goal is to provide a step-by-step account of the prototype's creation, emphasizing IoT technology and energy savings as the focus of our innovation.

The main components in our prototype's development were an SCT013-030 current sensor and an ESP32 dev kit V1 module. For the programming of the module, the Arduino IDE development environment was used, in which we configured the relevant libraries for the ESP32 module and Adafruit platform libraries (Adafruit MQTT Library) to facilitate the web publication of the results.

A. Material Description

- 1) **ESP32:** The ESP32 module from Espressif Systems is a System on Chip (SoC) that incorporates a dual-core 32-bit Tensilica Xtensa LX6 microprocessor (Fig. 1). This microprocessor typically operates at 160 MHz, but it is capable of achieving a clock speed of up to 240 MHz. The module integrates both a Wi-Fi communication stack and a Bluetooth Low Energy (BLE 4.1) communication stack, underlining its capacity for seamless integration into IoT applications.

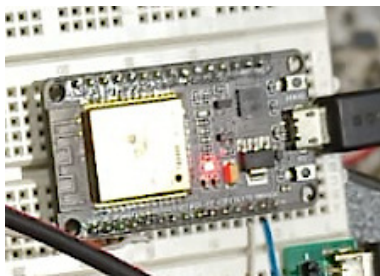


Fig. 1. Development board Dev kit V1 ESP32.

The programming of the ESP32 is achieved using the Arduino IDE and the C language, reinforcing its accessibility to developers across varying skill levels. The module employs a Reduced Instruction Set Computer (RISC) architecture, making it optimal for executing instructions at a high speed, thereby supporting more efficient energy use.

Featuring 30 pins, the module is flexible and adaptable for various inputs and outputs. The power supply voltage is 5 Vdc via the micro USB port or the Vin port, while all input and output pins operate at 3.3 Vdc. With 25 digital pins, the board can connect with a range of devices including sensors, LEDs, buttons, and other peripherals.

In addition, the ESP32 module includes two 12-bit ADC converters with 18 channels, expanding its interfacing capabilities. At an approximate cost of \$8 USD, the ESP32 provides a cost-effective solution for developing IoT devices with an emphasis on energy efficiency.

- 2) **SCT013-030 Current Sensor:** The SCT013-030 is a split-core current transformer typically used to measure alternating current (Fig. 2). One of the major advantages of this sensor is that it does not require cutting of wires for operation, thereby enhancing the

safety of its usage. The sensor is relatively affordable, with a market price of approximately \$9 USD.



Fig. 2. SCT013-030 current sensor.

Key features of the SCT013-030 current sensor are outlined below:

- **Input current:** The sensor can accurately measure AC in the range of 0-30A with a 1Vac output. This broad range accommodates a variety of applications, underscoring the sensor's versatility.
- **Non-linearity:** It exhibits a non-linearity of $\pm 1\%$, implying that the output is a highly accurate representation of the input. This characteristic is crucial for precise control and measurement tasks.
- **Bandwidth:** With a bandwidth of 1000Hz, the sensor is capable of handling fast-changing current levels, making it suitable for monitoring harmonic distortion.
- **Resistance grade:** The sensor has a B-grade resistance level, indicating its ability to withstand moderate current flows without performance degradation.
- **Working temperature:** The sensor can operate efficiently in a wide temperature range, from -25°C to 70°C , which ensures its performance under diverse environmental conditions.
- **Dielectric strength:** The sensor exhibits a dielectric strength of 1000Vac/1 min 5mA between its shell and output. This feature implies a high level of insulation, reducing the risk of electric shock.
- **Cable length and Size:** The sensor comes with a 1-meter long cable and has a compact size of 13mm x 13mm. This makes it convenient to integrate the sensor into various system configurations.

B. Signal Conditioning for the SCT013 Sensor

- 1) **Voltage signal conditioning:** At the outset of this project, the operating parameters of the current sensor were duly

identified, verifying its ability to measure up to the 11th harmonic, courtesy of its 1 kHz bandwidth. The signal fed into the current sensor or current transformer is an AC signal ranging from 0 to 30 A ac. The sensor then transforms this into a voltage level between 0-1 V_{rms} due to the presence of an internal burden resistor.

If the incoming current signal is in a sinusoidal mode, the output of the current transformer is a sinusoidal signal with an amplitude between 0 and 1 V_{rms}. Consequently, the maximum peak-to-peak voltage this signal can achieve is given by Eq. (1):

$$V_{pp} = V_{rms} \times 2\sqrt{2} = 2.82842 \quad (1)$$

Considering that the Analog to Digital Converter (ADC) of the ESP32 module only receives positive values between 0 and the reference voltage (in this case 3.3 VDC), it is necessary to condition the signal from the current sensor to fit within this operational window (Fig. 3). To achieve this, we decided to add a DC voltage of 1.56 V to the sensor's input signal. Therefore, the maximum and minimum peak voltages at the ADC input will be Eq. (2) and (3):

$$V_{pmax} = 1.56V_{dc} + \sqrt{2} = 2.97V_p \quad (2)$$

and

$$V_{pmin} = 1.56V_{dc} - \sqrt{2} = 0.1458V_p \quad (3)$$

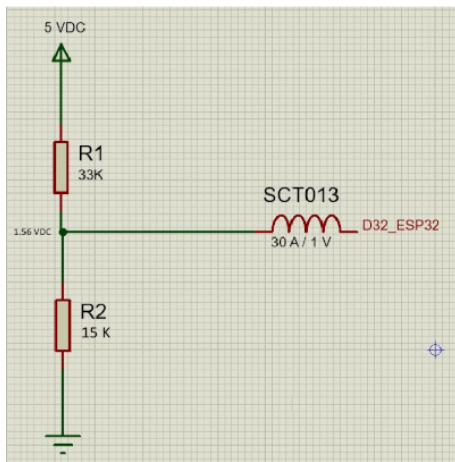


Fig. 3. Voltage divider on the analog input of the ESP32.

Through this voltage signal conditioning process, we ensure that the ADC of the ESP32 module operates within its specified range, optimizing the accuracy of our current measurements and contributing to the overall energy efficiency of the prototype.

C. Characterization Test for the SCT 013-030 Sensor Curve

To validate if the SCT013-030 sensor aligns with the manufacturer's linearity specifications, a characterization test was conducted. A resistive load was used alongside a stepwise variable voltage source that incremented by 0.5 V until an output current of 27A was achieved. This test's purpose was

twofold: firstly, to ensure the performance of the sensor under various operating conditions, and secondly, to formulate an associated equation reflecting the sensor's response, which would be instrumental in the prototype's implementation.

The test's resultant curve depicting the sensor's behavior is presented in the Fig. 4. It should be noted that this characterization is crucial not only for validation against the manufacturer's specifications but also as an input to the development of the control algorithm that drives the IoT device's energy savings.

The equation that describes the sensor's behavior, derived from the curve, is crucial in determining the precise measurements of current. This data is used to calibrate the signal conditioning process and to feed accurate current values into the system's control algorithm, ensuring the IoT device operates at optimal energy efficiency.

D. Software Development

The development of the firmware for the current meter encompassed several stages, each contributing to the comprehensive functionality of the IoT device. The process began with the generation of a vector comprised of 1024 twelve-bit values, corresponding to the samples of the current signal emanating from the sensor. This data acquisition is essential in capturing the intricate details of the current waveform and provides a robust base for the subsequent Fast Fourier Transform (FFT) algorithm implementation Eq. (4).

$$\text{Vectorcurrent} = [I_1, I_2, \dots, I_{1024}] \quad (4)$$

where I_i are the twelve-bit samples of the current signal.

After data collection, the selection and testing of the FFT algorithm were carried out. FFT is a powerful computational tool used to transform the acquired time-domain current samples into the frequency domain. The frequency domain representation of the current signal provides a more detailed understanding of the signal components, which aids in accurate and efficient current monitoring Eq. (5).

$$\text{FFT}(\text{Vector}_{\text{current}}) = [\text{Amp}_1, \text{Amp}_2, \dots, \text{Amp}_{1024}] \quad (5)$$

Subsequent to the FFT process, the final stage involved integrating the firmware with the Adafruit.io platform for data publication. This stage is of paramount importance in the IoT application, as it provides a way to share, visualize, and analyze the data generated by the IoT device in a user-friendly manner. This step bridges the gap between raw data collection and actionable insights, contributing significantly to the overall energy savings facilitated by the IoT device.

E. ADC Configuration and Sampling Time

The Analog-to-Digital Converter (ADC) was configured with a resolution of 12 bits and an internal reference voltage of 3.3 VDC, the default setting. This provided an operating voltage window from 0 to 3.3 VDC, corresponding to output values ranging from 0 to 4096. For ease of operation, a mapping function, `mapf()`, was implemented to convert the ADC output values to input values expressed in millivolts. The

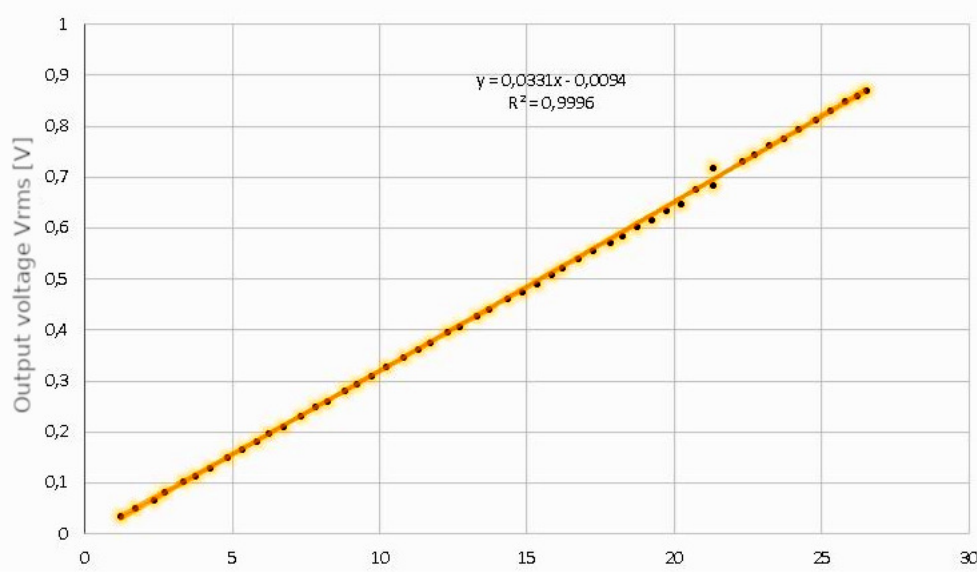


Fig. 4. Experimental characterization curve sensor SCT 013-030.

`mapf()` function accepts the maximum and minimum values of the ADC as parameters and returns a floating-point value between 0 and 3300 mV Eq. (6).

$$\begin{aligned} \text{Mapped Value (mV)} = \\ \text{mapf}(\text{ADC Value, ADC Minimum,} \\ \text{ADC Maximum, 0, 3300}) \quad (6) \end{aligned}$$

To sample the current waveform accurately, it was determined that data should be collected over ten periods of the fundamental grid frequency. Given that the grid frequency for Colombia is 60 Hz, this corresponds to a sampling period of approximately 166.66 ms. The sampling time and frequency can therefore be calculated as Eq. (7) and (8):

$$t_s = \frac{0.166666}{1024} = 0.00016275 \text{ s} \quad (7)$$

and

$$f_s = \frac{1}{t_s} = \frac{1}{0.00016275} = 6144 \text{ Hz} \quad (8)$$

This sampling frequency is more than sufficient for measuring up to the 15th harmonic (900 Hz), ensuring that the device captures a detailed picture of the current waveform for efficient energy monitoring and control.

F. Implementation of Fast Fourier Transform Algorithm

We utilized the Fast Fourier Transform (FFT) computation library authored by Robin Scheibler, which offers a comprehensive description on [FFT On The ESP32](#). This particular library algorithmically decomposes a discrete signal into its spectral constituents through a Radix-2 Decimation

in Frequency method. It employs the Bit-reversal technique to reorder the output vector, starting from the DC component (zero frequency at position zero of the output array) up to the sampling frequency divided by two for a unilateral transform or up to the sampling frequency for a bilateral transform.

The output vector of the algorithm contains complex values corresponding to each frequency in the measurement range. Hence, post-FFT, the amplitude calculations for each frequency are performed on the input vector. The fundamental frequency and its corresponding amplitude are identified as the maximum of the computed magnitudes Eq. (9).

$$A_f = \sqrt{\text{Real}^2 + \text{Imag}^2} \quad (9)$$

Here, A_f denotes the magnitude at a specific frequency, and this value should coincide with the amplitude at that frequency. Based on our practical observations, we found a need to correct these values. Thus, we introduced a correction factor for the amplitude of each frequency component, which we denoted as $I_{error} = 0.662946429$. This value was derived empirically using laboratory measurements and curve fitting techniques Eq. (10).

$$A_f = 0.662946429 \sqrt{\text{Real}^2 + \text{Imag}^2} \quad (10)$$

These corrective steps ensured the precision of the signal decomposition, thereby enabling more accurate energy consumption measurements and contributing to the broader goal of enhancing energy efficiency through IoT technologies.

G. Calculation of Power and Energy

The SCT013 sensor's current measurement range (0 to 30A) is mapped once the amplitude of the fundamental frequency is obtained. This mapping is accomplished through the function `Calculate_Irms()`. Assuming a sinusoidal input

voltage regime of 120 Vrms amplitude, the instantaneous power P_0 in kilowatts is defined as follows Eq. (11):

$$P_0 = \frac{V_{rms} \times I_{rms}}{1000} \quad (11)$$

The energy calculation used a base time of 1.04 seconds, the time it takes for the program to complete all computations, from sampling to THDi calculation, six times over. Hence, energy, E , is calculated as Eq. (12):

$$E = P(W) \times T_{(1040)ms}[h] = P(W) \times 1040[ms] \times \frac{1h}{3600 \times 10^3ms} \quad (12)$$

Given that energy accumulation should not exceed 24 hours, the accumulation timer and energy counter reset every 24 hours (86400 seconds).

H. Calculation of Current Harmonic Distortion

Once the FFT of the current signal and the corresponding harmonic amplitudes are obtained, the Total Harmonic Distortion of Current (THDi), caused by the presence of nonlinear loads in the system, is determined Eq. (13):

$$THD_i = \frac{\sqrt{\sum_{n=2}^{n_{max}} I_n^2}}{I_1} = \frac{I_H}{I_1} \quad (13)$$

Here, I_H denotes the root mean square (rms) value of the harmonic current, and I_1 represents the rms value of the fundamental current. This calculation offers an important insight into the system's performance, specifically concerning the influence of nonlinear loads, which directly impacts energy consumption and system stability. Thus, it provides a key factor for implementing energy-saving solutions in IoT environments (Fig. 5).

I. Prototype Testing

In order to comprehensively evaluate the performance of the prototype, a specialized testing setup was designed and implemented. This process required a current source capable of supplying current to a biphasic load. In this instance, the load was represented by four infrared lamps, typically used in automotive paint-drying ovens. This testbed, under normal operation, is capable of delivering 380V line-to-line (VLL). However, the load can withstand up to 220VLL.

The testing process started from a zero-baseline voltage, which was progressively increased. Simultaneously, the current values were measured using the SCT013 current sensor, an Extech clamp meter, and an AEMC 8220 power quality analyzer. This procedure was done to cross-verify and confirm the accuracy of the measurements taken by the developed prototype. The Fig. 6 illustrates the high voltage laboratory setup, which was utilized for performing these tests on the prototype.

Fig. 7 provides a comprehensive illustration of the intricate connection or measurement schema adopted in this study. At the core of this configuration are the three principal devices, which are pivotal to the research:

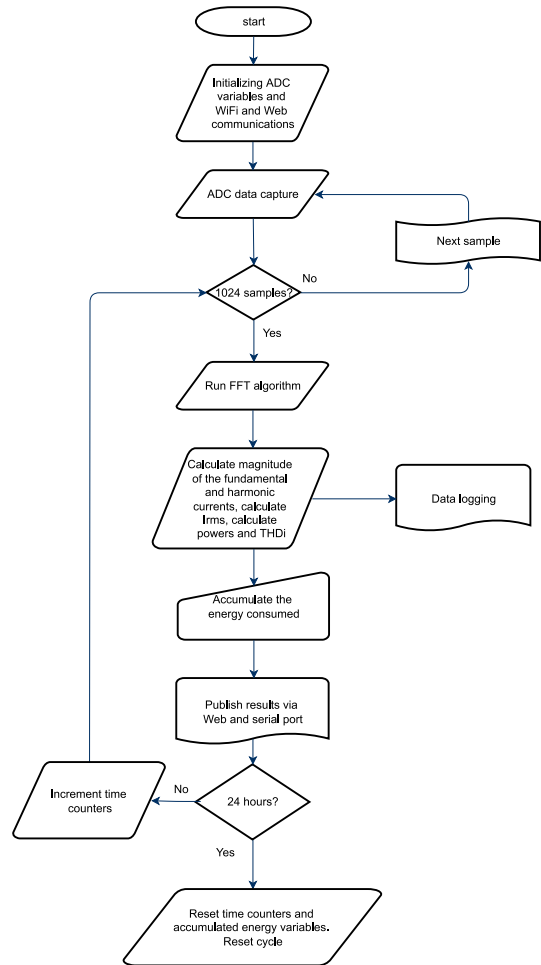


Fig. 5. Flowchart of the processing algorithm.

- **SCT013 Current Sensor:** This sensor is designed to measure the current flowing through a conductor without the need to interrupt the circuit. It uses a magnetic core to detect changes in current and convert it into a voltage, which can then be read by our prototype. Its non-invasive nature allows for safer and more convenient monitoring, especially in scenarios where continuous power supply is paramount.
- **Extech Clamp Meter:** Acting as a supplementary tool, the Extech clamp meter aids in the measurement process by clamping onto the conductor to determine current values. Its utility is evident when one wishes to validate readings quickly without delving into intricate circuit connections, providing a quick yet reliable snapshot of the current scenario.
- **AEMC 8220 Power Quality Analyzer:** This device is considered the gold standard in our research. The AEMC 8220 is a versatile tool capable of measuring multiple parameters, including voltage, current, and power quality attributes. In our schema, it serves the dual purpose of providing reference measurements and validating the accuracy of our prototype's readings. By comparing results from the IoT-based energy sav-



Fig. 6. Test bench.

ing prototype with the AEMC 8220, we ensure the credibility and reliability of our device.

The strategic arrangement of these devices is essential to achieve two key objectives: firstly, to enable accurate and consistent data acquisition by our IoT-based energy saving prototype; and secondly, to ensure that the measurements from the prototype can be cross-verified against established and well-regarded instruments in the industry, thereby reinforcing the validity and reliability of our findings.

A comprehensive view of the entire setup employed for prototype testing with the infrared lamps is depicted in Fig. 8. This illustration provides a clear visual of the prototype, connected and functioning within the wider measurement system.

Moreover, as shown in Fig. 7, the infrared lamps can be seen reaching incandescence during the testing phase. This phenomenon is a result of the power supplied to the lamps, showcasing the operational flow of the overall system. The lamps glowing red-hot emphasize the real-time nature of the prototype testing, hinting at the authenticity and practical applicability of this IoT-focused energy-saving system.

During the initial phase of testing, depicted in Fig. 9, the

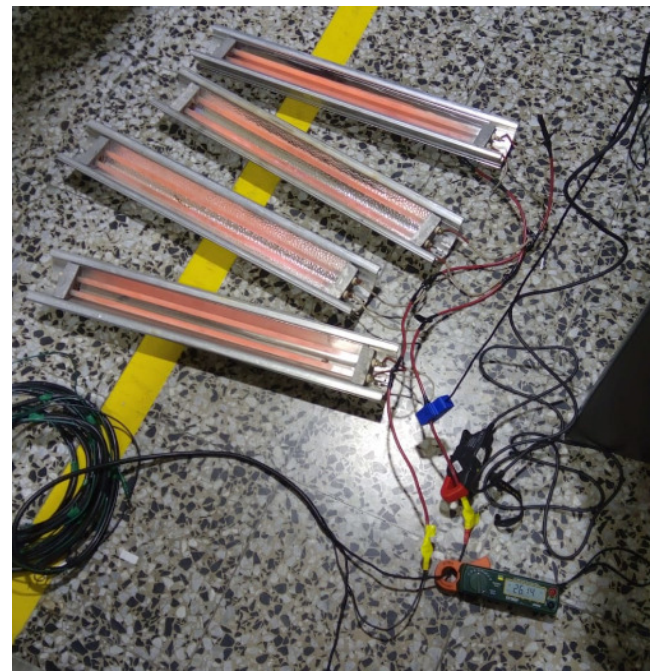


Fig. 7. Connection and operation of load and measuring instruments.

prototype exhibited behavior that deviated from the anticipated performance due to a confluence of factors. These variables are inherent in any experimental setup and include the accuracy of resistors, the measurement error introduced by the SCT013 current sensor, and the intrinsic variations in data acquired over time.

To account for these factors and to validate our prototype's performance, we compared the measurements obtained from our IoT-based energy-saving system with those acquired from two robust measurement tools: the Extech clamp meter and the AEMC 8220 power quality analyzer. This comparative approach allowed us to understand the performance nuances of the prototype and adjust the system to minimize the influence of external variables on the overall performance.

Following the preliminary testing phase and the insight gleaned from it, we made software adjustments to correct the discrepancies detected in the prototype's behavior. Our goal was to reduce the measurement error and improve the accuracy of the energy metrics provided by our IoT-based system.

During the subsequent testing phase, illustrated comprehensively in Fig. 10, we observed marked enhancements in the performance of our prototype. This phase of testing was crucial, as it allowed us to evaluate the modifications made after our initial tests.

- **Data Points Alignment:** The data points captured from our system were strikingly more aligned with those obtained from our reference measurement tools. Instead of a broad scatter or deviation, the points clustered around the readings from the AEMC 8220 and the Extech clamp meter. This closer alignment served as an early indication of the success of our recent tweaks and calibrations.



Fig. 8. Measurement and capture instruments during testing.

- **Reference Measurement Tools:** The choice of AEMC 8220 and Extech clamp meter as our reference tools was strategic. The AEMC 8220, being a renowned power quality analyzer, provided us with accurate and industry-accepted measurements. On the other hand, the Extech clamp meter gave us rapid and reliable snapshots of the current, serving as an essential tool for instant validation. Our prototype's readings moving closer to these reference tools' measurements was a significant achievement.
- **Software Adjustments:** The software adjustments we incorporated after the initial test phase were pivotal. These adjustments, a combination of algorithm tweaks and calibration methods, aimed at refining the measurement accuracy and eliminating any observed anomalies. The data from the subsequent testing phase strongly suggested that these software interventions played a major role in enhancing the prototype's performance.

Conclusively, the dataset from this testing phase clearly indicated that our prototype's behavior was now in tight congruence with the expected outcomes, grounded on the reference tools' readings. This not only solidified our confidence in the software changes we had implemented but also underscored the prototype's potential for reliable energy measurements in real-world applications.

IV. RESULTS

Examining the response portrayed in the various figure plots, it is evident that the devised energy meter's performance measures favorably with existing industry standard devices. Despite the intricate characteristics inherent in this prototype, it not only fulfils the capabilities of measuring active power and Total Harmonic Distortion of current (THDi) but accomplishes these tasks with significant cost advantages.

These characteristics are particularly vital for efficient energy management, and by having a solution that can provide such capabilities at a much lower cost, we are propelling ourselves towards improved energy savings. The insight from the experimental results suggest that integrating IoT technology with traditional electronics and control engineering principles can indeed form a basis for a high-performing, yet cost-effective solution for power and THDi measurement.

The overall performance of our device, as suggested by the curves depicted in the figures, confirms the robustness of the design and the success of our software adjustments in refining the meter's precision. These results provide substantial evidence that our approach to a more economical and compact energy meter can be invaluable in power quality and energy management systems where budget and space constraints are of importance.

While the prototype exhibited excellent performance with a precision rate of 91% compared to the AEMC 8220, it's worth noting that in scenarios with fluctuating loads, the device consistently measured active power within a margin of 2%. Moreover, the THDi measurements provided insights into the harmonics present, showcasing its capability to be a tool not just for energy measurement but also for preliminary power quality analysis in households. The low cost of approximately \$24 USD makes this device particularly attractive for residential settings, especially in developing regions where budget constraints are paramount.

A. Limitations

Despite the evident success and promise of our IoT energy meter, it is crucial to acknowledge certain limitations of our study:

- **Scope of Testing:** The comparative testing was primarily performed against the AEMC 8220 energy meter. While it provides a benchmark, the results might differ when compared with a broader range of energy meters available in the market.
- **Prototype Stage:** The device is still in its prototype stage. Real-world application and longevity tests are required to ensure its robustness and durability over extended periods.

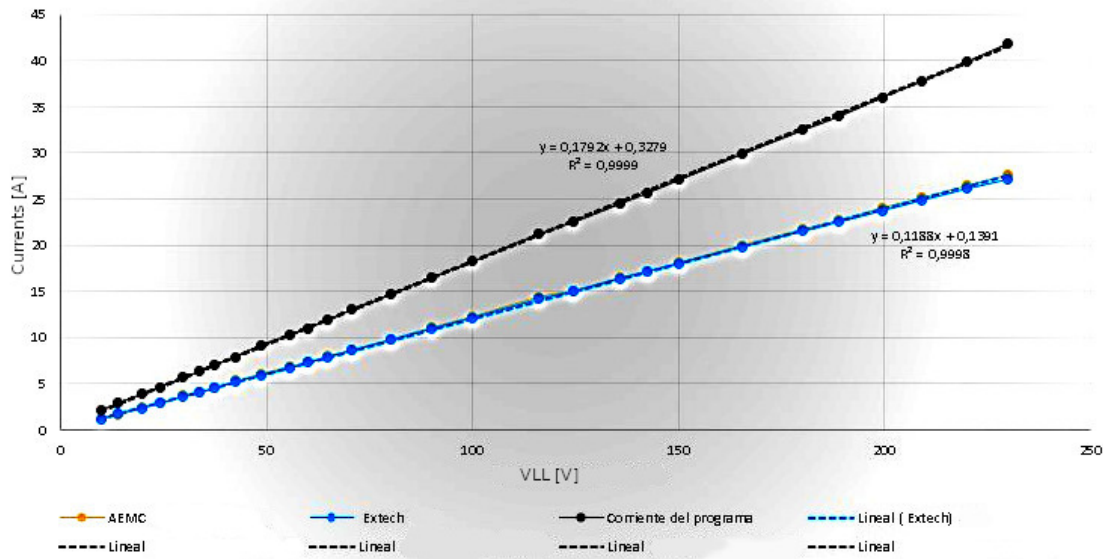


Fig. 9. Initial prototype testing.

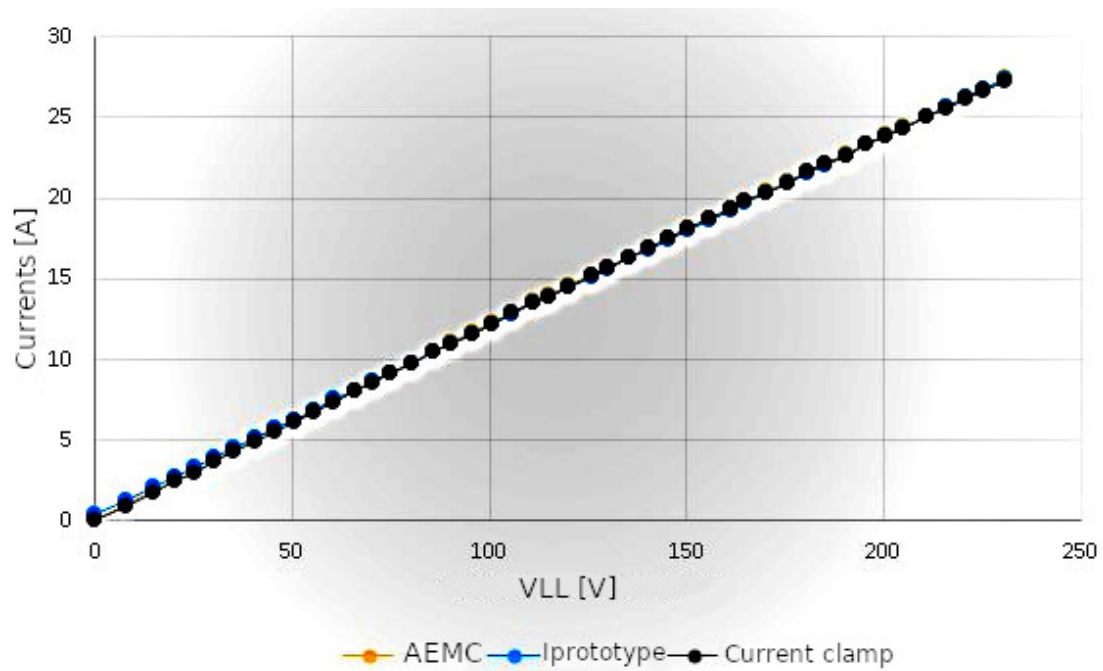


Fig. 10. Software-adjusted prototype testing.

- **Potential Interferences:** The prototype was tested in controlled environments. External factors like electromagnetic interferences or extreme environmental conditions, which can potentially affect the performance, were not extensively studied in this research.
- **Software Limitations:** While open-source software offers cost advantages and customizability, it might not be as optimized or stable as proprietary software solutions in certain scenarios. This might lead to performance or stability issues in some edge cases.

V. CONCLUSION

Throughout this study, we focused on the design and development of a cost-effective, user-friendly, and feature-rich IoT energy meter as an alternative to other market offerings. Our goal was to enable mass production of the meter to aid in critical global endeavors, including energy efficiency and optimization, especially considering the growing prevalence of nonlinear loads in residential settings.

The successful development of a prototype has been achieved using accessible and cost-effective materials, in con-

junction with open-source software, costing approximately \$24 USD. Comparative testing against a well-regarded AEMC 8220 energy meter yielded a precision of 91% in relation to key parameters such as active power and Total Harmonic Distortion of current (THDi).

The developed prototype's low cost and high quality offer substantial benefits to the realm of energy efficiency. It equips end-users with a simple, attainable tool for monitoring and managing their household energy consumption, thereby advancing residential energy optimization.

As we look to the future, it is feasible to extend the functionality of the device to include bidirectional energy measurement capabilities. This would enable the support of alternative energy sources for low-income users. Furthermore, the inclusion of THD measurements can assist in calculating the impact of harmonics on the distribution network, leading to improved power quality management at a residential level.

VI. FUTURE RESEARCH DIRECTIONS

While the current research and development have proven successful in designing an economical and precise IoT energy meter, there are several promising avenues for future research and enhancement of the device:

- **Integration with Renewable Energy Sources:** As the global shift towards green energy continues, integrating the energy meter with renewable energy sources such as solar and wind can be invaluable. It would be insightful to research how our meter could be enhanced to support not just traditional power sources but also renewable ones, offering users a comprehensive view of their energy consumption and generation.
- **Advanced Harmonics Analysis:** With the increasing use of nonlinear devices in homes, harmonics play a crucial role in power quality. Future research can delve deeper into advanced harmonic analysis techniques and provide users with detailed reports and insights into their energy consumption patterns, enabling them to make informed decisions on managing and reducing their harmonic footprint.
- **Machine Learning and Predictive Analysis:** By incorporating machine learning algorithms, the energy meter can offer predictive analytics on energy consumption, allowing users to anticipate their energy needs and adjust accordingly. This research could lead to the development of an intelligent system that not only monitors but also predicts and optimizes energy consumption based on historical data and user behavior.

ACKNOWLEDGMENTS

This work was supported by the Universidad Distrital Francisco José de Caldas, specifically by the Technological Faculty. The views expressed in this paper are not necessarily endorsed by Universidad Distrital. The authors thank all the students and researchers of the research group ARMOS for their support in the development of this work.

REFERENCES

- [1] S. Chaudhari, P. Rathod, A. Shaikh, D. Vora, and J. Ahir, "Smart energy meter using arduino and gsm," in *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017, pp. 598–601.
- [2] Z. Sultan, Y. Jiang, A. Malik, and S. F. Ahmed, "Gsm based smart wireless controlled digital energy meter," in *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 2019, pp. 1–6.
- [3] Q. Sun, H. Li, Z. Ma, C. Wang, J. Campillo, Q. Zhang, F. Wallin, and J. Guo, "A comprehensive review of smart energy meters in intelligent energy networks," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 464–479, 2016.
- [4] S. Gowrishankar, N. Madhu, and T. G. Basavaraju, "Role of ble in proximity based automation of iot: A practical approach," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2015, pp. 400–405.
- [5] P. A. Berde and R. G. Bhavani, "Investigation for reducing energy consumption for a university campus in dubai using automation," in *2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, 2015, pp. 358–363.
- [6] T. Xia, C. Liu, X. Zheng, M. Lei, S. Wang, and Y. Wang, "Research on field testing method of digital energy meters based on digital reference meter with high accuracy," in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*, 2018, pp. 1–9.
- [7] N. L. Andrei, V. Tanasiev, M. Sanduleac, and A. Badea, "Smart metering platform as a solution for data analysis," in *2017 International Conference on ENERGY and ENVIRONMENT (CIEM)*, 2017, pp. 495–499.
- [8] E. Trujillo Lorenzo, "Analizador de consumo de potencia eléctrica con arduino," 2015.
- [9] C. Fuertes, "Diseño e implementación de un módulo que permita la generación de armónicos y su corrección para el análisis y monitoreo con un analizador virtual de red," 2018.
- [10] R. T. Mathew, S. Thattat, K. Anirudh, V. P. K. Adithya, and G. Prasad, "Intelligent energy meter with home automation," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, 2018, pp. 1–4.
- [11] F. H. Martínez Sarmiento, "El fenómeno de distorsión armónica en redes eléctricas," *Tecnura*, vol. 5, no. 9, p. 46–54, Jul. 2001.
- [12] U. Ramani, S. kumar, T. Santhoshkumar, and M. Thilagaraj, "Iot based energy management for smart home," in *2019 2nd International Conference on Power and Embedded Drive Control (ICPEDC)*, 2019, pp. 533–536.
- [13] M. Babiuch, P. Folytynek, and P. Smutny, "Using the ESP32 microcontroller for data processing," in *2019 20th International Carpathian Control Conference (ICCC)*. IEEE, 2019.
- [14] J. A. Aguirre-Nunez, L. M. Garcia-Barajas, J. de Jesus Hetnandez-Gomez, J. P. Serrano-Rubio, and R. Herrera-Guzman, "Energy monitoring consumption at IoT-edge," in *2019 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. IEEE, 2019.
- [15] H. K. Patel, T. Mody, and A. Goyal, "Arduino based smart energy meter using gsm," in *2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 2019, pp. 1–6.
- [16] Y. Ochoa, J. Rodriguez, and F. Martinez, "Low cost regulation and load control system for low power wind turbine," *Contemporary Engineering Sciences*, vol. 10, no. 28, pp. 1391–1399, 2017.
- [17] P. K. Steimer, "Power electronics, a key technology for energy efficiency and renewables," in *2008 IEEE Energy 2030 Conference*, 2008, pp. 1–5.
- [18] E. P. de Medellín ESP Unidad Centro de Excelencia Técnica Normalización y Laboratorios, "Nt -13 norma técnica: Calidad de potencia de redes de distribución," EPM, Tech. Rep., 2019.
- [19] A. García Granados, "Diseño y desarrollo de un sistema de monitorización remoto de parámetros de la red eléctrica, basado en web y herramientas open source," 2020.
- [20] R. Sheeba, N. Naufal, S. Nadera Beevi, A. R. Nair, S. Amal, A. S. Kumar, A. Mohan, P. Arunjith, U. Aswin, T. Aswanth, J. Joseph, and J. Jose, "Real-time monitoring of energy meters using cloud storage," in *2021 IEEE International Power and Renewable Energy Conference (IPRECON)*, 2021, pp. 1–5.

- [21] M. Binti, A. Ibrahim, and N. Hasnati, "Energy meter (e-mon) with iot monitoring (web & apps)," *Journal of Technical and Vocational Education*, vol. 1, no. 1, pp. 128–137, 2019.
- [22] B. Suleimenov, K. Iskakov, and D. Nartova, "An efficient digital hardware-software complex for electricity metering," in *2022 IEEE 23rd International Conference of Young Professionals in Electron Devices and Materials (EDM)*. IEEE, 2022.
- [23] M. M. Martín-Lopo, J. Boal, and Á. Sánchez-Mirallas, "A literature review of IoT energy platforms aimed at end users," *Computer Networks*, vol. 171, no. 1, p. 107101, 2020.

A Novel Convolutional Neural Network Architecture for Pollen-Bearing Honeybee Recognition

Thi-Nhung Le¹, Thi-Minh-Thuy Le², Thi-Thu-Hong Phan^{3*}, Huu-Du Nguyen⁴, Thi-Lan Le⁵

Faculty of Information Technology, Vietnam National University of Agriculture, Hanoi, Vietnam^{1,2}

School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam^{1,5}

Department of Artificial Intelligence, FPT University, Danang, Vietnam³

School of Applied Mathematics and Informatics, Hanoi University of Sciences and Technology, Hanoi, Vietnam⁴

Abstract—Monitoring the pollen foraging behavior of honeybees is an important task that is beneficial to beekeepers, allowing them to understand the health status of their honeybee colonies. To perform this task, monitoring systems should have the ability to automatically recognize images of pollen-bearing honeybees extracted from videos recorded at the beehive entrance. In this paper, a novel convolutional neural network architecture is proposed for recognizing pollen-bearing and non-pollen-bearing honeybees from their images. The performance of the proposed model is illustrated based on a real dataset and the obtained results show that it performs better than some other state-of-the-art deep learning architectures like VGG16, VGG19, or Resnet50 in terms of both accuracy and execution time. Thus, the proposed model can be considered an effective algorithm for designing automatic honeybee colony monitoring systems.

Keywords—Pollen-bearing honeybee; image classification; convolutional neural network; honeybee monitoring system; pollen dataset

I. INTRODUCTION

Honeybees bring great benefits to human life. Products from honeybees such as honey, propolis, and pollen bring high economic efficiency. They are both popular components for daily consumption and an important source of raw materials in the production of medicines and beauty care. The global honey market was valued at 8.9 billion dollars by the year 2022, and it is predicted to reach 12.6 billion dollars in the year 2030, according to a report in VANTAGEMarketResearch¹. In addition, honeybees are known as the most common and effective pollinators [1]. A large-scale survey in [2] has shown that honeybee pollination has contributed to increased yields and improved quality for many crops worldwide. For example, in the US, honeybees have increased fruit setting by 60% and seed yield by 20% for almonds; in Argentina, honeybees simultaneously increase fruit setting (by 15%) and the content of fruit sugar for apples, thereby increasing profits by 70%; and in Brazil, honeybees increased soybean yield by 18.9%. Besides that, the pollination of honeybees also contributes to the preservation of the diversity of plant ecosystems.

To take care of honeybee colonies, beekeepers must monitor the health of the colonies regularly. This task requires gathering information about the activities, status, and behaviors of honeybees in the colony, including pollen foraging behavior. In fact, pollen is the leading food of honeybees. It

provides proteins, lipids, vitamins, and minerals necessary for the growth and reproduction of honeybees [3]. Information about the foraging behavior of honeybees can bring valuable understanding about the pollen source status in the habitat, the need for food, the increase of individuals, and the health of the whole honeybee colony. As a result, it allows beekeepers to understand the status of their honeybee colonies and detect unusual problems in the colonies for timely intervention. Recognizing honeybees bringing pollen back to the hive is then an effective solution for monitoring the beehives.

In recent years, thanks to the application of IoT (Internet of Things) technologies, several automatic honeybee monitoring systems have been deployed. These systems use surveillance cameras to record the activities of honeybees at the beehive entrance, then use different techniques to extract and analyze information from the recorded images [4]. Due to its powerful ability in image data processing, the Convolutional Neural Network (CNN) is perhaps the most widely used technique in these systems. In the context of recognizing pollen-bearing honeybee images, many CNN-based models have been designed. For example, several well-known CNN-based models like VGG16, VGG19, Resnet50, and DarkNet53 have been applied in [5] towards precise recognition of pollen-bearing honeybees. Rodriguez et al. [6] tested with several types of CNN architectures and showed that shallow-CNN architecture gives higher recognition accuracy than machine learning methods such as SVM (Support Vector Machine), Naive Bayes, or K-nearest neighbors. The authors also provided a real dataset of pollen-bearing and non-pollen-bearing honeybee images. However, we have found that there are a few mislabeled samples in this dataset where some images of non-pollen-bearing honeybees were assigned as pollen-bearing honeybees and vice versa. In addition, the use of complex structures for these CNN-based models requires a large number of samples for training and testing, leading to a significant cost of calculation and resources for the execution. To provide an effective model for designing automatic beehive monitoring systems, in this study, we propose a novel CNN architecture for recognizing pollen-bearing honeybee images. Here, we aim to find a model that is better than existing models in terms of both efficiency and execution cost. The performance of the proposed model is validated by comparing it with several other complex models like VGG16, VGG19 [7], and Resnet50 [8] architectures using the same dataset.

In summary, the main contributions of the study are as follows:

¹<https://www.vantagemarketresearch.com/industry-report/honey-market-2138>

- We adjust a dataset published in a previous study by assigning correct labels to some previously mislabeled samples.
- We propose a novel structure for a CNN-based model to classify the images of pollen-bearing and non-pollen-bearing honeybees. We also suggest a data augmentation technique to handle the cases where there are few observed samples.
- We verify the efficacy and cost of the proposed method by extensive experiments, achieving an absolute accuracy for pollen-bearing honeybee recognition on the testing set.

The rest of the paper is organized as follows. Section II is to present the recent related works in the literature. In Section III, we describe the dataset considered in the study and the proposed CNN architecture. The experiments and the obtained results are presented in Section IV. Finally, Section V is for some concluding remarks.

II. RELATED WORK

In this section, we briefly discuss recent studies related to the classification models of pollen-bearing and non-pollen-bearing honeybees.

For pollen-bearing honeybee image recognition, many studies relied on image processing techniques and conventional machine learning algorithms [9]. Babic et al. [10] applied background subtraction using a Mixture of Gaussian for the segmentation of honeybees. Then, based on the difference in color variance and eccentricity features between pollen-bearing and non-pollen-bearing honeybees, the authors used the Nearest Mean Classifier to classify them. The accuracy achieved by the classifier is 88.7%. However, the classification accuracy depends on the results of background subtraction and the light source in the video recording area. Yang and Collins [11] used color thresholding and the Mixture of Gaussian to detect and extract images of individual honeybees in frames captured from video recorded at the beehive entrance and track them using Kalman filter and Hungarian algorithm. The bee blob analysis method from the binary image of each frame was applied to remove the main body of the honeybee and retain only the pollen blobs. The two main features of the pollen blobs, including the area of the pollen blobs and the location of them relative to the bee's body, will be used to remove noise blobs. Finally, the pollen sacs detection results are combined with the previous honeybee detection and tracking model to identify if a honeybee bears pollen sacs. The test results with several videos show that the pollen measurement model has the highest sensitivity of 76%. In [12], the authors conducted experiments using two methods to segment honeybee images: the k-means algorithm and the algorithm that only considers the b component of the CIE LAB color space. Then, the SVM classifier with Gaussian kernel was used to classify the pollen-bearing and non-pollen-bearing honeybee images based on the Dense SIFT (Dense Scale Invariant Feature Transform) descriptors and the VLAD (Vector of Locally Aggregated Descriptors) encoder. The test results show that the method that combines the k-means segmentation algorithm and the classifier based on the descriptors on the decorrelated channels

gives the highest value of the area under the ROC curve (AUC-ROC) at 0.915. In [6], authors performed the classification of pollen-bearing and non-pollen-bearing honeybees with three traditional methods: K-Nearest Neighbor algorithm, Naive Bayes statistical algorithm, and Support Vector Machines with linear and non-linear kernel functions. The results show that the SVM RBF method (Support Vector Machine with Radial Basis Function) with PCA (Principal Component Analysis) preprocessing technique and using the Gaussian feature map gives the highest accuracy at 91.16%.

In machine learning models, the important features were often selected from the inputs manually and subjectively. This can greatly affect the performance of the model once the key elements are not considered. To overcome this disadvantage, recent research suggest using models based on deep learning, more specifically, different CNN architectures. Rodriguez et al. [6] conducted experiments with 1-layer and 2-layer Shallow-CNN models, VGG16, VGG19, and Resnet50. The results show that all these models achieve high accuracy in which Shallow-CNN with small step size gives the highest accuracy of 96.4%, followed by VGG19, VGG16, and Resnet50 with an accuracy of 90.2%, 87.2%, 61.7%, respectively. Sledevič [13] investigated different CNN architectures with different numbers of hidden layers. After several experiments, the author stated that the architecture consisting of three hidden layers 7-7, 5-5, and 3-3 is the most suitable for classifying pollen-bearing and non-pollen-bearing honeybees, achieving a 94% accuracy. In [14], a pollen sac detection model on an individual honeybee image is used to classify a honeybee image as a pollen-bearing or non-pollen-bearing honeybee. This detection model uses Faster R-CNN architecture with the core for classification as VGG16. When a pollen sac is detected on an individual honeybee image, it is marked by a bounding box labeled "pollen" and a numerical value that is the confidence score of the detection. When the confidence score is greater than or equal to a predefined threshold, it is counted as a pollen sac, and the individual honeybee image is counted as a pollen-bearing honeybee image. This model has a pollen detection accuracy of 81.5%. Ngo et al. [15] relied on the YOLOv3-tiny model to detect and classify objects. Since YOLO-v3 treats the object classification problem as a regression problem, whereby an input image is divided into grid cells, each grid cell is responsible for detecting a target honeybee. This model allows the simultaneous detection of multiple objects on a frame belonging to one of two classes of pollen-bearing and non-pollen-bearing honeybees. The obtained results from this study showed that the proposed model gives a classification accuracy of 94%. In another research, nine different pre-trained CNN models, including VGG16, VGG19, Resnet50, ResNet101, Inception V2, Inception V3, Xception, DenseNet201, and DarkNet53 have been explored [5]. The authors also considered the influence of color by applying some image preprocessing techniques to the input dataset. The experimental results showed that the DarkNet53 and VGG16 architectures attained higher recognition accuracy than the others. In [16], the authors first tested the image classification using the transfer learning method with seven pre-trained Deep Neural Networks (DNNs) including AlexNet, DenseNet201, GoogLeNet, ResNet101, ResNet18, VGG16, and VGG19. After that, the authors continued to experiment with the SVM classifier using shallow features, deep features,

and shallow+deep features extracted from the DNNs. Three different standard datasets were used for the training and evaluation of models. The experimental results showed no significant difference in performance between them. For the pollen-bearing honeybee image dataset, the transfer learning method with pre-trained DNN yielded the highest accuracy of 99.07%.

III. MATERIALS AND METHODS

A. Data Description

In this study, the Pollen dataset published in [6] has been considered. The dataset is available and can be accessed publicly at GitHub². The authors stated that at the beginning, there were 810 honeybee images extracted and manually annotated from videos taken at the beehive entrance under natural light conditions. However, after being curated, many have been removed due to misclassification. Thus, the final data downloaded from the source above contains only 714 images. The photo was built by fixing the size of the cropping rectangle to 180×300 pixels, containing a fully visible image of a single honeybee. The images are also adjusted to make sure that the honeybees are facing upward in all images. Each photo was then labeled with pollen or non-pollen. Out of 714 images, 369 are labeled as pollen (P) and the rest 345 are non-pollen (NP). Fig. 1 presents some images of pollen-bearing and non-pollen-bearing honeybees in this dataset.



Fig. 1. Images of pollen-bearing honeybees (a) and non-pollen-bearing honeybees (b).

Based on this dataset, we have conducted several experiments to investigate the performance of the proposed model. The obtained results have shown that our model has misidentified some images from pollen-bearing honeybees to non-pollen-bearing ones and vice versa. We then carried out a thorough analysis of these misidentified images and found that some of them were mislabeled. According to our knowledge, images *NP24865-145r* and *NP27452-203r* are images of honeybees that bear pollen but are annotated as non-pollen-bearing ones (NP). Meanwhile, images *P7660-97r*, *P7776-99r*, *P11440-32r*, and *P11762-35r* in the dataset are images of honeybees that do not bear pollen but are annotated as pollen-bearing honeybees (P). Fig. 2 shows these mislabeled images.

TABLE I. THE STRUCTURE OF THE POLLEN DATASET

Dataset	Class label	Number of images
Original	Pollen	369
	Non-pollen	345
Corrected	Pollen	367
	Non-pollen	347

We have relabeled these images and used the corrected dataset to train and test the performance of the proposed model (as well as control models). The structure of the original Pollen dataset and the relabeled one is summarized in Table I.

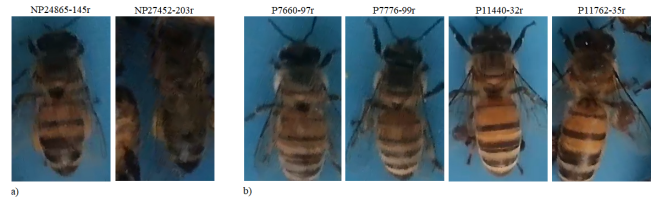


Fig. 2. Mislabeled images in the Pollen dataset, from Pollen to Non-Pollen (a) and vice versa (b).

B. Proposed Method

1) *Convolutional neural networks*: Convolutional Neural Networks (CNNs) refer to a well-known deep learning algorithm specialized in handling image data. The basic architecture of a CNN model consists of three main types of layers, as displayed in Fig. 3.

The functions of each type of layer are as follows:

- The first type of layer of a basic CNN architecture is Convolution, the core of a CNN model used to extract various features from the input. The mathematical convolutional operation is performed in this layer, between the input and a filter. The dot product is taken between the filter and the parts of the input by sliding the filter over the image. The output from each layer containing information about the image like corners and edges is then fed to the next layer to learn other input features.
- Following the convolutional layers are the Pooling layers. These layers summarize the features extracted from the previous convolution layers, aiming to decrease the size of the obtained feature map and reduce computational costs. Several types of pooling operations can be used in a CNN model depending on the specific situation, such as Max Pooling and Average Pooling.
- The last Fully connected layers perform the classification task based on the features extracted from previous layers, mapping the representation between the input and the output. They generate scores for each class, then use them for the final classification.

In general, there is no universal optimal model for all datasets. For different problems, one should design different models with different structures to achieve the best performance. Choosing the right model is the key to solving many

²<https://github.com/piperod/PollenDataset>

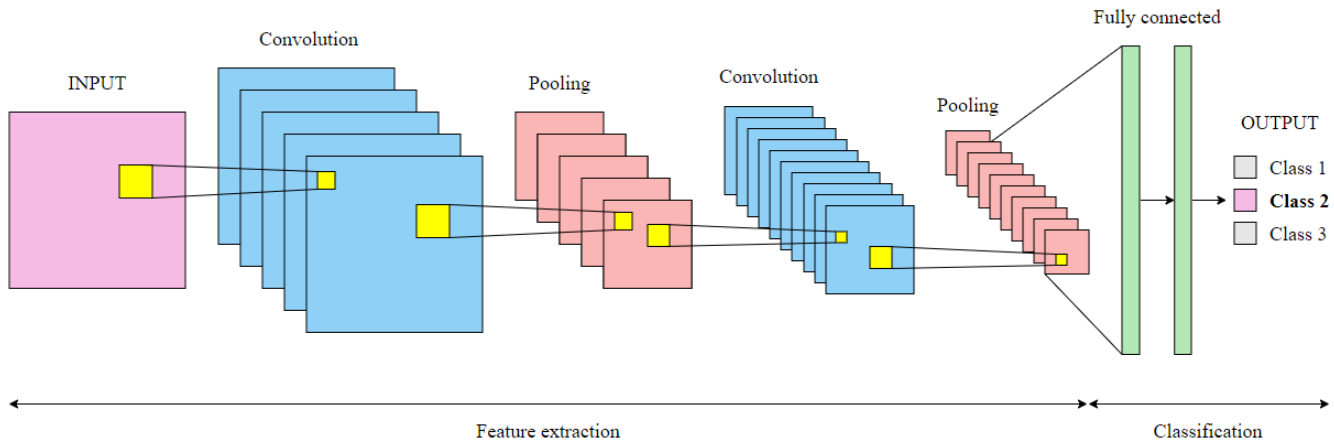


Fig. 3. Architecture of a basic CNN model.

problems in practice. More introduction and discussion about the architectures of CNN-based deep learning model can be seen in [17] and [18].

2) *The proposed CNN model:* As discussed above, the CNN architectures have been widely applied to the problem of recognizing pollen-bearing honeybees, some of them achieved up to about 99% accuracy. Several transfer learning models whereby different CNN architectures such as VGG16, VGG19, Resnet50, and Resnet101, etc. are pre-trained and some SVM classification models are based on features extracted from CNN architectures, were presented in [5] and [16]. However, the use of these models requires significant costs due to their complex architectures. To overcome this problem and aim for simplicity and efficiency in use, we thought about using a basic CNN model. This idea can be verified by investigating the basic CNN structures with different hyperparameters. In particular, using the Grid search method, we have figured out an optimal architecture for the CNN model in classifying pollen-bearing and non-pollen-bearing honeybees. The proposed architecture comprises:

- 4 convolutional layers equipped with a ReLU (Rectified Linear Unit) activation function,
- 5 max-pooling layers,
- 1 flatten layer, a dense layer with a ReLU activation function, and a dense layer with a Sigmoid activation function.

The use of this simple architecture obviously makes the model lighter than other pre-trained deep-learning models like VGG16, VGG19, and Resnet50 which contain more layers and parameters. In addition, in this study, we use some data augmentation techniques such as image rescaling, random rotating, shifting (horizontally and vertically), shearing, random zooming, random flipping, and nearest filling. These techniques enrich the data by generating different variations from the original images which will be used in different epochs of the model training process, thereby improving the performance of the classification model. Fig. 4 shows a visualization of several images obtained after applying data augmentation techniques to a honeybee image.

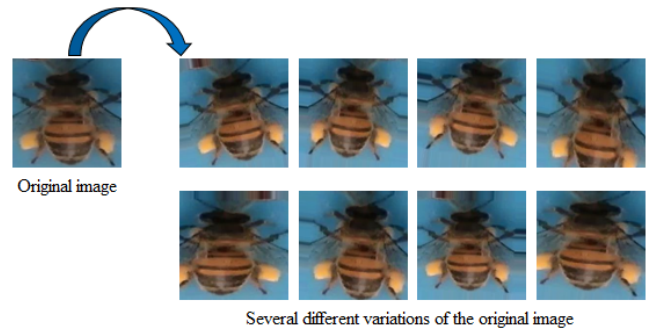


Fig. 4. An example of data augmentation.

The architecture of the proposed model is illustrated in Fig. 5. Each input is a 224×224 RGB image containing the image of an individual honeybee. After passing through the convolutional layers and the max-pooling layers to extract the important features, it is fed to fully connected layers. A predefined threshold is used to classify whether the honeybee image is a pollen-bearing honeybee or not. The performance of the proposed method will be discussed in the sequel.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

In this study, two schemes of splitting the Pollen dataset are considered. By the first scheme, as in several previous studies, we randomly divide the corrected Pollen dataset into three subsets, i.e., the training set, the validation set, and the testing set at a ratio of 6:1:3. Accordingly, 60% of the samples corresponding to 428 images (which include 221 images of pollen-bearing honeybees and 207 images of non-pollen-bearing honeybees) are for model training, 10% of the samples corresponding to 70 images (which include 36 images of pollen-bearing honeybees and 34 images of non-pollen-bearing honeybees) are for model validation, while the remaining 60% of the samples corresponding to 216 images (which include 110 images of pollen-bearing honeybees and 106 images of non-pollen-bearing honeybees) are for model testing. Moreover, to investigate the effect of partitioning data

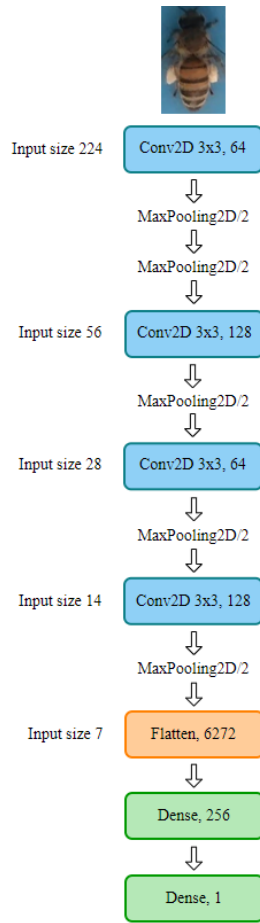


Fig. 5. The proposed CNN architecture.

TABLE II. TWO SCHEMES OF SPLITTING THE POLLEN DATASET

Scheme	Class label	Training set	Validation set	Testing set
1	Pollen	221	36	110
	Non-pollen	207	34	106
2	Pollen	183	36	148
	Non-pollen	173	34	140

into training, validation, and testing sets on the model performance, we design the second scheme where the Pollen dataset is split into the three subsets at a ratio of 5:1:4, namely fewer samples for training and more samples for testing compared to the first scheme. The details of the number of images in each subset of each scheme are presented in Table II.

The experiments were performed on Google Colab, using Python 3 Google Compute Engine backend (GPU) with a system RAM of 83.5 GB, GPU RAM of 40 GB, and Disk of 166.8 GB.

After hyperparameters tuning, hyperparameters are set up for model training as follows: batch size is 4, the number of epochs is 25, the initial learning rate is 0.001, and the optimizer is Adam.

B. Evaluation Metrics

To evaluate the performance of the proposed pollen-bearing honeybee recognizing model comprehensively, in this study,

Precision, Recall, F1-score, Accuracy, Loss, and AUC-ROC metrics have been utilized. These are widely used metrics to assess the performance of the classification models.

- Precision, Recall, F1-score, and Accuracy are computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-score} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where

- TP: number of pollen-bearing honeybees images that are properly classified as pollen-bearing honeybees images;
- FP: number of non-pollen-bearing honeybees images that are misclassified as pollen-bearing honeybees images;
- FN: number of pollen-bearing honeybees images that are misclassified as non-pollen-bearing honeybees images;
- TN: number of non-pollen-bearing honeybees images that are properly classified as non-pollen-bearing honeybees images.

- Loss (Binary Cross Entropy) is calculated as follows:

$$\text{Loss} = \frac{-1}{N} \sum_{i=1}^N (y_i \log_e p_i + (1 - y_i) \log_e (1 - p_i)) \quad (5)$$

where

- N is the number of images;
- y_i is the real label of the i th image ($y_i = 1$ if the i -th image is a pollen-bearing honeybee image; $y_i = 0$ if the i -th image is a non-pollen-bearing honeybee image);
- p_i is the probability of the event predicting the i -th image as a pollen-bearing honeybee image ($1 - p_i$ is the probability of the event predicting the i -th image as a non-pollen-bearing honeybee image).

- AUC-ROC: one of the most important evaluation metrics for checking any classification model's performance is calculated as the area under the ROC (Receiver Operating Characteristics) curve.

From the above definitions, the larger the values of Precision, Recall, F1-score, Accuracy, and AUC-ROC, and the smaller the value of Loss, the better the model is at classifying classes in a dataset.

TABLE III. THE PERFORMANCE OF CNN-BASED MODELS ON THE FIRST SCHEME OF SPLITTING THE POLLEN DATASET

Method	Precision	Recall	F1-score	Accuracy	Loss	AUC-ROC
VGG16 ([6])	-	-	-	87.20	-	-
VGG16 ([5])	-	-	-	94.80	-	-
VGG16 (ours)	98.67	98.58	98.61	98.61	10.86	0.9994
VGG19 ([6])	-	-	-	90.20	-	-
VGG19 ([5])	-	-	-	98.20	-	-
VGG19 (ours)	96.48	96.24	96.29	96.30	14.99	0.9961
Resnet50 ([6])	-	-	-	61.70	-	-
Resnet50 ([5])	-	-	-	86.60	-	-
Resnet50 (ours)	99.53	99.55	99.54	99.54	4.91	0.9955
Proposed CNN	100.00	100.00	100.00	100.00	1.32	1.0000

TABLE IV. THE EXECUTION TIME OF CNN-BASED MODELS ON THE FIRST SCHEME OF SPLITTING THE POLLEN DATASET

Method	Training time (s)	Testing time (s)
VGG16	272.830	133
VGG19	265.359	167
Resnet50	7471.634	39
Proposed CNN	150.543	6

C. Experimental Results and Discussion

The performance of the proposed method and the corresponding execution time on the first scheme of splitting the Pollen dataset is presented in Table III and Table IV. For the purpose of comparison, we also show the performance and the execution time of other CNN-based transfer learning methods using the same dataset in the literature.

Several important remarks can be drawn from these tables as follows.

- The proposed CNN model provides the best performance in terms of all the metrics. Although the use of other CNN-based models results in quite an impressive efficiency (for instance, an accuracy of 99.54% with Resnet50, and 98.61% with VGG16), our proposed method can still achieve higher performance, with an absolute efficiency of 100% for the metrics of Precision, Recall, F1-score, and Accuracy, and the maximum value is 1 for the metric of AUC-ROC. It also leads to the smallest value of the Loss of 1.32. This means the proposed model can accurately recognize all pollen-bearing and non-pollen-bearing honeybee images in the Pollen dataset.
- After correcting the mislabeled images, the accuracy of other CNN-based models is generally improved. For example, based on the original Pollen dataset, the Resnet50 model in [6] and [5] provided an accuracy of 61.70% and 86.60%, respectively. Meanwhile, on the corrected dataset, it can reach an accuracy of 99.54%.
- Thanks to its simple architecture with fewer layers than some other CNN architectures such as VGG16, VGG19, and Resnet50, the proposed model also reduces significantly the execution time for both training and testing processes. Indeed, it took only 150.543 seconds for training and 6 seconds for testing. Meanwhile, the second-fastest models asked for about 265.359 seconds for training (VGG19) and 39 seconds for testing (Resnet50), which are significantly slower than the proposed model, as can be seen in Table IV. This finding has a practical meaning as it allows the

designing of efficient real-time recognition systems of pollen-bearing honeybees.

Fig. 6 and Fig. 7 show the curves of the training and validation accuracy, and the training and validation loss of the models compared. As can be seen from these figures, the proposed CNN model gives a higher performance than the other models. This is accordant with the results discussed above.

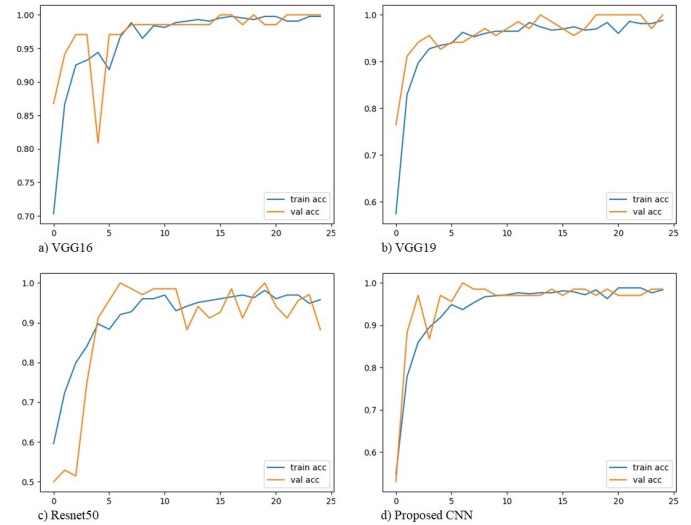


Fig. 6. Training and validation accuracy curve.

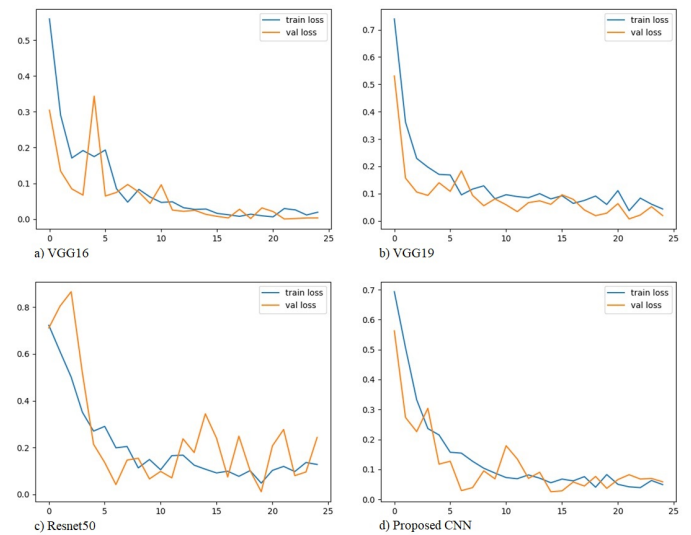


Fig. 7. Training and validation loss curve.

In Tables V and VI, we present the experimental results obtained from using the second scheme of splitting the Pollen dataset at the ratio 5:1:4. Since the scheme has not been considered in previous studies, we present the performance of our experiments only. The same result as the first scheme can also be witnessed in these two tables where our proposed method still brings the best Precision, Recall, F1-score, Accuracy, and Loss in the fastest processing time. However, the performance of all models, in this case, has been reduced a bit compared

TABLE V. THE PERFORMANCE OF CNN-BASED MODELS ON THE SECOND SCHEME OF SPLITTING THE POLLEN DATASET

Method	Precision	Recall	F1-score	Accuracy	Loss	AUC-ROC
VGG16	98.27	98.25	98.26	98.26	9.24	0.9961
VGG19	97.61	97.54	97.57	97.57	13.03	0.9941
Resnet50	95.73	95.40	95.47	95.49	9.58	0.9966
Proposed CNN	98.95	98.99	98.96	98.96	8.02	0.9921

TABLE VI. THE EXECUTION TIME OF CNN-BASED MODELS ON THE SECOND SCHEME OF SPLITTING THE POLLEN DATASET

Method	Training time (s)	Testing time (s)
VGG16	217.666	179
VGG19	241.568	223
Resnet50	6295.647	49
Proposed CNN	123.874	10

to the first scheme. For example, the proposed model does not achieve an absolute accuracy as in the first scheme, instead, it decreases to 98.96%. This result can be explained by the reduced number of samples in the training set. As a result, the model learns less from the training set, resulting in reduced classification performance.

V. CONCLUSION

In this paper, we have proposed a novel convolutional neural network model for classifying pollen-bearing and non-pollen-bearing honeybee images. Rather than using complex and pre-trained CNN models, we design a basic CNN architecture with a few layers, leading to a lighter and also more efficient model. We have also corrected some mislabeled samples from a widely used dataset in the literature. The performance of the proposed CNN model has been investigated and compared with other models based on this corrected dataset. The obtained results have shown that our method leads to the best performance in terms of both accuracy and execution time. In particular, it could identify correctly 100% all the pollen-bearing and non-pollen-bearing honeybee images from the testing set in the shortest time.

There are still several limitations that should be considered before deploying the use of the proposed model in designing automated systems to recognize pollen-bearing honeybees in practice. For example, the efficiency of the model was verified based on a small dataset that contains 714 images only. Its performance should be validated on other datasets with larger sizes. In addition, the choice of hyperparameters of the proposed architecture is suitable for the current dataset, but may not be for other datasets. Therefore, it would be better to have another method to find hyperparameters that are optimal for each dataset. From this point of view, some optimization algorithms such as Random search or Bayesian optimization could be applied for future work. In addition, the model can be applied to process honeybee images for some other related tasks, like counting the number of pollen-bearing honeybees (for the purpose of measuring the amount of pollen carried by honeybees to the hive), classifying pollen, or recognizing disease-carrying honeybees. However, its performance needs to be verified for each specific situation.

ACKNOWLEDGMENT

The authors would like to thank the Vietnam National University of Agriculture for sponsoring this research through the University-level science and technology project with the title "Research on deep learning algorithms of convolutional neural network and their applications in recognizing pollen-bearing honeybee images" and number "T2022-10-39".

REFERENCES

- [1] U. Joshi, K. Kothiyal, Y. Kumar, and R. Bhatt, "Role of honeybees in horticultural crop productivity enhancement," *International Journal of Agricultural Sciences*, vol. 17, no. AAEBSDD, pp. 314–320, 2021.
- [2] S. A. M. Khalifa, E. H. Elshafey, A. A. Shetaia, A. A. A. El-Wahed, A. F. Algethami, S. G. Musharraf, M. F. AlAjmi, C. Zhao, S. H. D. Masry, M. M. Abdel-Daim, M. F. Halabi, G. Kai, Y. A. Naggari, M. Bishr, M. A. M. Diab, and H. R. El-Seedi, "Overview of bee pollination and its economic value for crop production," *Insects*, vol. 12, no. 8, 2021.
- [3] K. A. Stoner, H. P. Hendriksma, and S. Tosi, "Pollen as food for bees: Diversity, nutrition, and contamination," *Frontiers in Sustainable Food Systems*, vol. 6, no. 1129358, 2023.
- [4] I. F. Rodriguez, J. Chan, M. Alvarez Rios, K. Branson, J. L. Agosto-Rivera, T. Giray, and R. Mègret, "Automated video monitoring of unmarked and marked honey bees at the hive entrance," *Frontiers in Computer Science*, vol. 3, no. 769338, 2022.
- [5] F. C. Monteiro, C. M. Pinto, and J. Rufino, "Towards precise recognition of pollen bearing bees by convolutional neural networks," *In Iberoamerican Congress on Pattern Recognition*, pp. 217–226, 2021, May.
- [6] I. F. Rodriguez, R. Megret, E. Acuna, J. L. Agosto-Rivera, and T. Giray, "Recognition of pollen-bearing bees from video using convolutional neural network," in *In 2018 IEEE winter conference on applications of computer vision (WACV)*, 2018, March, pp. 314–322.
- [7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [9] S. Bilik, O. Bostik, L. Kratochvila, A. Ligocki, M. Poncak, T. Zemcik, M. Richter, I. Janakova, H. P., and K. Horak, "Machine learning and computer vision techniques in bee monitoring applications," *arXiv preprint arXiv:2208.00085*, 2022.
- [10] Z. Babic, R. Pilipovic, V. Risojevic, and G. Mirjanic, "Pollen bearing honey bee detection in hive entrance video recorded by remote embedded system for pollination monitoring," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3, pp. 51–57, 2016.
- [11] C. Yang and J. Collins, "Improvement of honey bee tracking on 2d video with hough transform and kalman filter," *Journal of Signal Processing Systems*, vol. 90, pp. 1639–1650, 2018.
- [12] V. Stojnić, V. Risojević, and R. Pilipović, "Detection of pollen bearing honey bees in hive entrance images," in *In 2018 17th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 2018, March, pp. 1–4.
- [13] T. Sledevič, "The application of convolutional neural network for pollen bearing bee classification," in *In 2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 2018, November, pp. 1–4.
- [14] C. Yang and J. Collins, "Deep learning for pollen sac detection and measurement on honeybee monitoring video," in *In 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, 2019, December, pp. 1–6.
- [15] T. N. Ngo, D. J. A. Rustia, E. C. Yang, and T. T. Lin, "Automated monitoring and analyses of honey bee pollen foraging behavior using a deep learning-based imaging system," *Computers and Electronics in Agriculture*, vol. 187, no. 106239, 2021.
- [16] S. K. Berkaya, E. S. Gunal, and S. Gunal, "Deep learning-based classification models for beehive monitoring," *Ecological Informatics*, vol. 64, no. 101353, 2021.

- [17] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
- [18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

Tomato Disease Recognition: Advancing Accuracy Through Xception and Bilinear Pooling Fusion

Hoang-Tu Vo, Nhon Nguyen Thien, Kheo Chau Mui
Software Engineering Department, FPT University, Cantho City, Vietnam

Abstract—Accurate detection and classification of tomato diseases are essential for effective disease management and maintaining agricultural productivity. This paper presents a novel approach to tomato disease recognition that combines Xception, a pre-trained convolutional neural network (CNN), with bilinear pooling to advance accuracy. The proposed model consists of two parallel Xception-based CNNs that independently process input tomato images. Bilinear pooling is applied to combine the feature maps generated by the two CNNs, capturing intricate interactions between different image regions. This fusion of Xception and bilinear pooling results in a comprehensive representation of tomato diseases, leading to improved recognition performance. Extensive experiments were conducted on a diverse dataset of annotated tomato disease images to evaluate the effectiveness of the suggested approach. The model achieved a remarkable test accuracy of 98.7%, surpassing conventional CNN approaches. This high accuracy demonstrates the efficacy of the integrated Xception and bilinear pooling model in accurately identifying and classifying tomato diseases. The implications of this research are significant for automated tomato disease recognition systems, enabling timely and precise disease diagnosis. The model's exceptional accuracy empowers farmers and agricultural practitioners to implement targeted disease management strategies, minimizing crop losses and optimizing yields.

Keywords—Tomato disease recognition; Xception; Bilinear pooling; convolutional neural networks; disease management

I. INTRODUCTION

Tomato plants are a vital crop worldwide, serving as a staple in numerous cuisines and contributing to global food security. Besides, regular consumption of tomatoes can contribute to improved health and a decreased susceptibility to various ailments, including cancer, osteoporosis, and cardiovascular disease. Those who incorporate tomatoes into their diet on a consistent basis are found to have a lower likelihood of developing cancer, encompassing lung, prostate, stomach, cervical, breast, oral, colorectal, esophageal, pancreatic, and several other forms of cancer [1]. However, tomato harvests are at risk of a variety of diseases that can cause significant yield losses and quality deterioration. Early and reliable identification of such diseases is necessary for implementing timely disease management strategies and minimizing agricultural losses. Tomato diseases are a significant concern for tomato growers worldwide, including Vietnam [2], [3], [4].

Techniques derived from Machine Learning (ML) and Deep Learning (DL) have been extensively utilized in image recognition across diverse domains, including agriculture [5], [6], medicine [7], [8], self-driving cars [9], [10], etc. Numerous studies have explored the application of these technologies to achieve accurate image recognition in these respective fields.

The purpose of this paper is to address the crucial need for accurate classification of tomato diseases in order to effectively manage them and maintain agricultural productivity. The study presents a novel approach to tomato disease recognition by combining Xception, a pre-trained convolutional neural network (CNN), with bilinear pooling to enhance accuracy. Extensive experiments were conducted on a diverse dataset of annotated tomato disease images to evaluate the effectiveness of the suggested approach. The model obtained an impressive test accuracy of 98.7%.

The paper is structured as follows: Section II offers an extensive literature review, presenting pertinent background information. Section III details the methodology utilized for Tomato Disease Recognition, including the Data Set, Data Preparation, and Model Evaluation Metrics. Section IV describes the experimental system and final results. Lastly, Section V concludes the study by summarizing the findings and providing concluding remarks. Section VI gives future directions of research.

II. RELATED WORKS

In recent years, improvements in DL techniques, particularly convolutional neural networks (CNNs) and transfer learning models (TL), have shown promising results in automating disease recognition tasks in agriculture. Zahid Ullah et al. [11] presents a hybrid deep learning approach, EffiMobNet, combining EfficientNetB3 and MobileNet models with techniques to handle overfitting, achieving a 99.92% accuracy in accurately detecting tomato leaf diseases. This study [12] utilizes pre-trained CNNs, specifically Inception V3 and Inception ResNet V2, to classify healthy and unhealthy tomato leaf images, achieving high accuracy (99.22%) and low loss (0.03%) with dropout rates of 50% and 15%, respectively. Sachin Kumar et al. [13] utilized a dataset of 6,594 tomato leaves, including six disease classes and one healthy class, from Plant Village, and achieved a significant accuracy of 96.35% using the ResNet-50 model. The authors in [14] suggest a method for classifying tomato leaf diseases employing transfer learning and feature concatenation by leveraging pre-trained kernel from MobileNetV2 and NASNetMobile models. They extract features from these models, concatenate them, and then reduce the dimensionality using kernel principal component analysis. The effectiveness of the concatenated features is confirmed through experimental results, with multinomial logistic regression achieving the best performance among the evaluated traditional machine learning classifiers, achieving an average accuracy of 97%.

The authors in the article [15] focus on developing a combined model for identifying tomato diseases utilizing image

data. Seven architectures, including VGG16, ResNet50, and various EfficientNet models, are evaluated for performance using transfer learning. The best-performing models are then combined using a weighted average ensemble, resulting in a suggested model with an accuracy of 98.1%. Nagamani H S et al. in the study [16] explores the identification of diseases affecting tomato leaves using ML techniques, including Fuzzy-SVM, CNN, and Region-based Convolutional Neural Network (R-CNN). Various image processing and feature extraction methods are employed, and R-CNN achieves the greatest accuracy of 96.735% in classifying different disease types. In this study [17], a DL model combining CNN and SVM is deployed to recognize and categorize tomato leaf images into 8 classes, including 7 prevalent diseases and a healthy class. The model is trained on a dataset of 8,000 photos and achieves an accuracy of 92.6% by utilizing CNN for feature extraction and SVM for classification.

Sanjeela Sagar et al. in this paper [18] presents an experimental study comparing traditional ML algorithms (RF, SVM, NB) with a deep learning CNN algorithm in order to classify tomato leaf disease. The results show that the CNN approach outperforms traditional methods, achieving over 95% accuracy in detection and classification. The authors in the article in this study [19] focus on using CNN methods, specifically the VGG model, for the detection of Multi-Crops Leaf Disease (MCLD). The trained model successfully classifies disease-affected leaves with high accuracy, achieving 98.40% accuracy for grapes and 95.71% accuracy for tomatoes. In this paper [20], the authors conduct a thorough evaluation of deep-learning approaches utilizing pre-trained CNN models and the PyTorch framework for classifying instances of diseases affecting tomato plants. Several models, including EfficientNet-B0, ResNext-50-32x4d, and MobileNet-V2, were tested, and ResNext-50-32x4d achieved the highest accuracy of 90.14%.

The paper in [21] presents an approach for classifying seven different types of tomato illnesses using DL models trained on a dataset of 10,448 images. The trained models demonstrated high accuracy, with the best testing precision reaching 95.71%. Sakkarvarthi, Gnanavel, et al in the article [22] proposes a deep-learning-based agricultural disease detection technique, employing a CNN approach in order to detect and classify diseases. The model, consisting of a pair of convolutional and pooling layers, exceeded the performance of the pre-trained InceptionV3, ResNet 152, and VGG19 models, achieving 98% training accuracy and 88.17% testing accuracy. In this paper [23], Singh, Rahul, et al. utilizes transfer learning with the EfficientNetB3 model for leaf classification, using a dataset of 11 different leaf types collected from an internet database. With a batch size of 32, the model is trained for 15 iterations and evaluated using the Adam optimizer, achieving an accuracy of 0.94. Sultana, Irene, et al. [24] present a dataset of 14,529 tomato leaf images containing ten different infections. InceptionV3 and ResNet-50 serve as employed learning algorithms, leveraging transfer learning techniques to train a classifier. The proposed deep learning model achieves promising results with an 85.52% accuracy rate for InceptionV3 and 95.41% for ResNet-50. The study in [25] aims for the purpose of identifying the presence of early blight infestation affecting tomato plants using a CNN approach. Various image processing techniques are applied to refine the dataset, and the CNN model is trained and evaluated

using different performance metrics, achieving a high accuracy of 98.10% with specific hyperparameters. The article [26] presents a CNN model that combines elements of different approaches for classifying diseases in tomato leaf images, utilizing well-known CNN architectures and feature transfer techniques. The suggested approach obtains high accuracy rates of 98.3% and 96.3% for both the dataset specifically designed for detecting tomato leaf diseases and the dataset collected in Taiwan, respectively.

The purpose of this study is to introduce an innovative method for tomato disease recognition by combining Xception, a pre-trained CNN, with bilinear pooling to enhance accuracy. The proposed model incorporates two parallel Xception-based CNNs that independently process tomato images and utilizes bilinear pooling to capture complex interactions between image regions. This fusion of Xception and bilinear pooling yields a comprehensive representation of tomato diseases, resulting in improved recognition performance.

III. METHODOLOGY

A. Data Collection and Preparation

In this research, a dataset comprising 32,535 images was used, obtained from the PlantVillage dataset [27] and Kaggle. The Tomato Disease dataset consists of 10 diseases and 1 healthy class, including Late blight (Class1), healthy (Class2), Early blight (Class3), Septoria leaf spot (Class4), Tomato Yellow Leaf Curl Virus (Class5), Bacterial spot (Class6), Target Spot (Class7), Tomato mosaic virus (Class8), Leaf Mold (Class9), Spider mites Two-spotted spider mite (Class10), and Powdery Mildew (Class11).

The examples of tomato disease pictures from the dataset are displayed in Fig. 1, and Fig. 2 illustrates the distribution of the dataset. Before training and evaluating the model, the images are preprocessed by resizing them to 224x224 and applying an image preprocessing function. The dataset is then split into 25,851 photos for the training dataset, 4,010 photos for the validation dataset, and 2,674 photos for the test dataset.

B. Proposed Model

This paper introduces a novel approach for accurate classification of tomato diseases using a model that combines Xception [28], a pre-trained CNN, with bilinear pooling. The proposed model consists of two parallel Xception-based CNNs that independently process tomato images, with bilinear pooling capturing intricate interactions between different regions of the images. The model aims to accurately classify images into 11 different classes.

The proposed model is based on the Xception architecture. It takes an input image of size 224x224x3. The model consists of two parallel Xception layers that process the input image independently, resulting in two sets of feature maps with dimensions 7x7x2048. These feature maps are then combined by the so-called bilinear pooling layer (by taking their outer product). An average pooling layer is applied to reduce the spatial dimensions to 1x1 while preserving the depth of 2048. The output is flattened to a 1D vector of size 2048.

Next, a batch normalization layer is used, followed by a 256-unit dense layer. A dropout layer is introduced to prevent



Fig. 1. Sample tomato disease in tomato disease dataset.

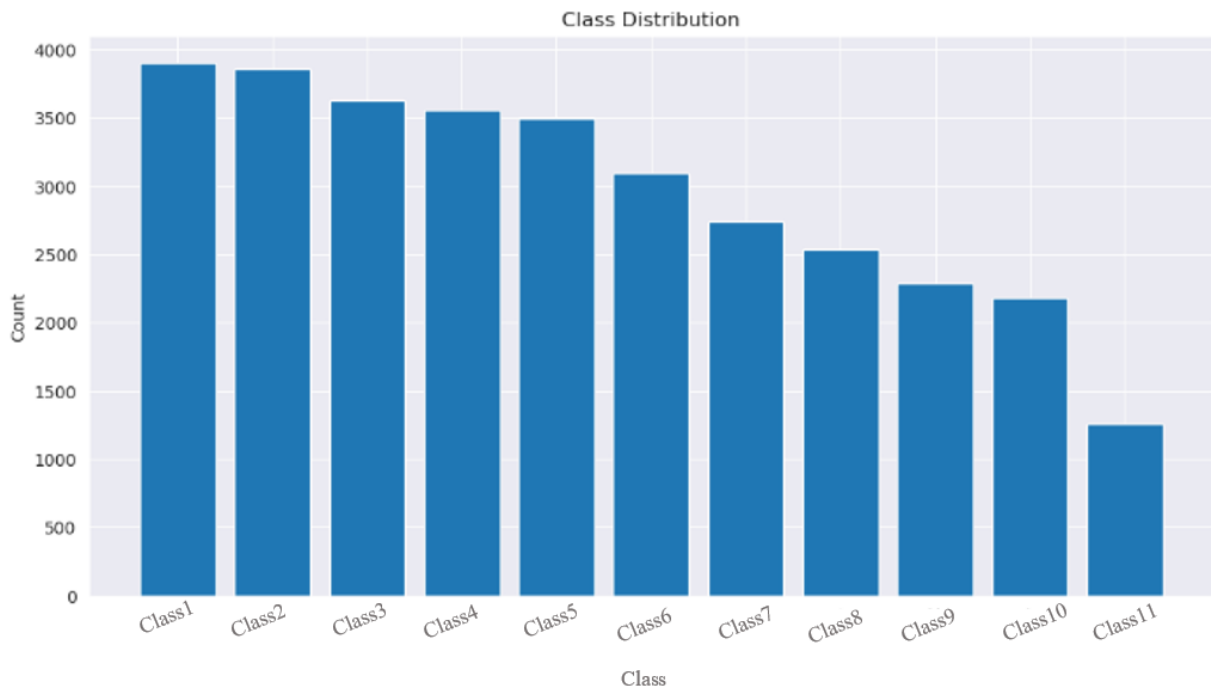


Fig. 2. A dataset distribution.

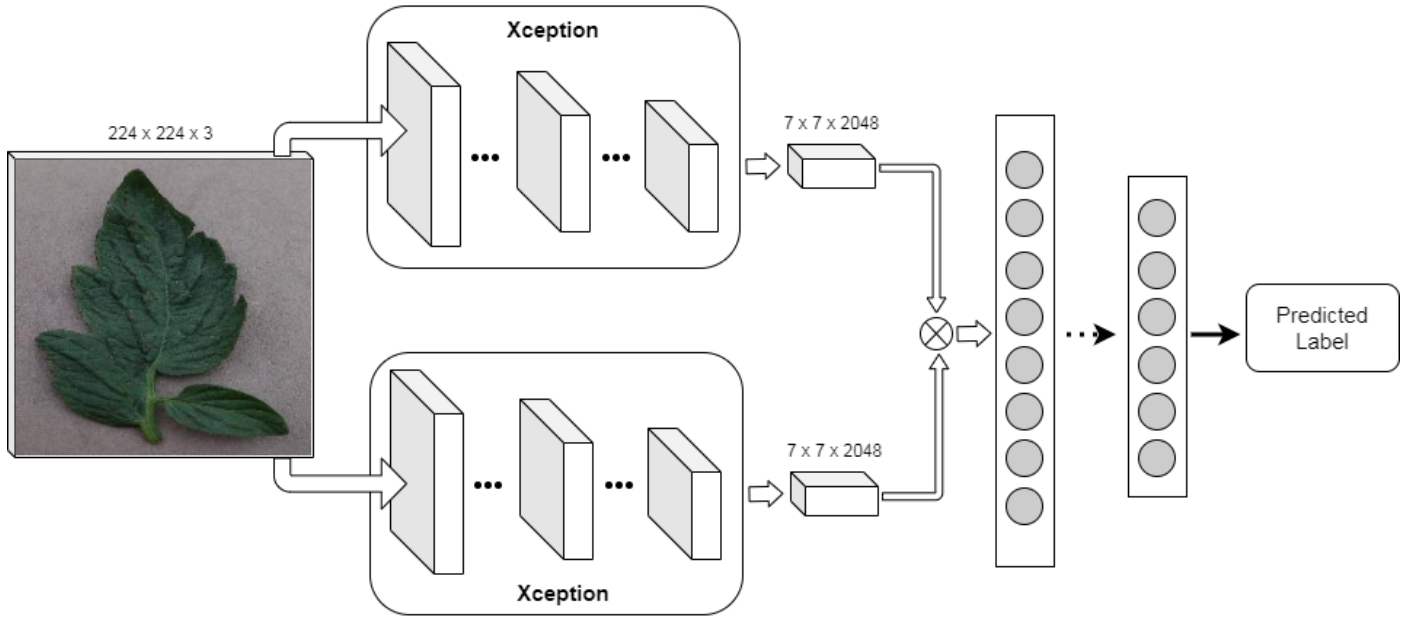


Fig. 3. The model proposed for classifying tomato diseases.

overfitting, and another dense layer with 11 units (matching the number of classes) is added as the final output layer. Fig. 3 shows the suggested model.

In the suggested model, a learning rate (LR) scheduling strategy [29] is integrated into the training phase. LR scheduling is a dynamic technique that adapts the LR, determining the size of the steps taken in the gradient descent optimization algorithm, during the model's training process. The objective of this approach is to improve the model's convergence by gradually decreasing the LR as training progresses.

In this specific model, the LR is reduced by a factor of 0.5 after two epochs, allowing for adjustments if training accuracy does not show improvement. To mitigate the risk of overfitting, the architecture of the model utilizes regularization techniques, which aid in avoiding overfitting and improving the overall performance of the model. These measures are crucial for real-world applications where optimal performance is of utmost importance.

The proposed model exhibits a substantial parameter count, with a total of 42,258,523 parameters. Among these, 42,145,371 parameters are trainable, meaning they are optimized and adjusted during the training process to enhance the model's performance. Additionally, there are 113,152 non-trainable parameters, which consist of fixed or pre-defined values that remain unchanged during training.

The suggested model and its underlying architecture can be visualized in Fig. 4, while Fig. 5 provides a comprehensive representation of important details. This includes information about the model's layers, the output shape, the trainable parameter count, and the overall number of trainable parameters that are examined for each layer in the model.

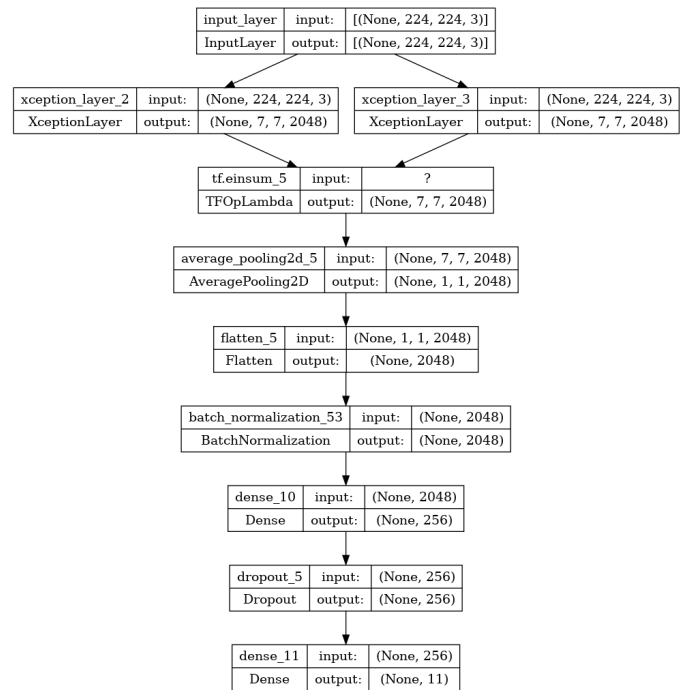


Fig. 4. The suggested model architecture proposed in this study.

C. Performance Evaluation Measures

The performance of the proposed integrated Xception and bilinear pooling model was comprehensively evaluated using a range of essential metrics. Precision, which quantifies the ratio of correctly predicted positive instances to the total

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	[(None, 224, 224, 3 0)]	0	[]
xception_layer_2 (XceptionLayer)	(None, 7, 7, 2048)	20861480	['input_layer[0][0]']
xception_layer_3 (XceptionLayer)	(None, 7, 7, 2048)	20861480	['input_layer[0][0]']
tf.einsum_5 (TFOPLambda)	(None, 7, 7, 2048)	0	['xception_layer_2[0][0]', 'xception_layer_3[0][0]']
average_pooling2d_5 (AveragePooling2D)	(None, 1, 1, 2048)	0	['tf.einsum_5[0][0]']
flatten_5 (Flatten)	(None, 2048)	0	['average_pooling2d_5[0][0]']
batch_normalization_53 (BatchNormalization)	(None, 2048)	8192	['flatten_5[0][0]']
dense_10 (Dense)	(None, 256)	524544	['batch_normalization_53[0][0]']
dropout_5 (Dropout)	(None, 256)	0	['dense_10[0][0]']
dense_11 (Dense)	(None, 11)	2827	['dropout_5[0][0]']

Total params: 42,258,523
Trainable params: 42,145,371
Non-trainable params: 113,152

Fig. 5. The layers of suggested model.

predicted positive instances, reflected the model’s ability to minimize false positives in classifying tomato diseases. Recall, capturing the ratio of correctly predicted positive instances to the actual total positive instances, demonstrated the model’s proficiency in identifying all relevant disease cases. The F1-score, a harmonic mean of precision and recall, provided a balanced assessment of the model’s precision-recall trade-off. Accuracy, a fundamental measure in classification tasks, gauges the proportion of correctly predicted instances out of the total instances in the dataset. In the context of tomato disease recognition, accuracy indicates the model’s overall correctness in identifying and classifying different disease types from input tomato images.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 - Score = \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

In which, TP represents True Positive, TN signifies True Negative, FP represents False Positive, and FN stands for False Negative.

IV. RESULTS

A. Environmental Settings

The experimental results were obtained by conducting the experiments on the Kaggle platform. The system used for the experiments had 13GB of RAM and a GPU Tesla P100-PCIE with 16GB of memory. The training of the model spanned across 30 epochs, and a batch size of 32 was used during the training process.

B. Evaluation Overall

The confusion matrix, showcasing the results of the proposed model, is presented in Fig. 6 and Fig. 7. Additionally, Fig. 8 and Fig. 9 illustrate the performance metrics, such as loss and accuracy, that were evaluated during both the model’s training and validation stages. The model achieves its highest accuracy at the 17th epoch and exhibits the lowest loss at the 28th epoch.

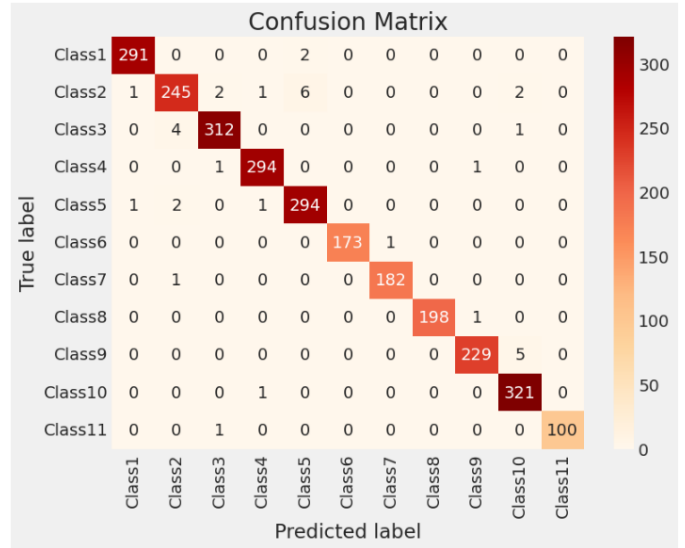


Fig. 6. Proposed model for tomato disease classification.

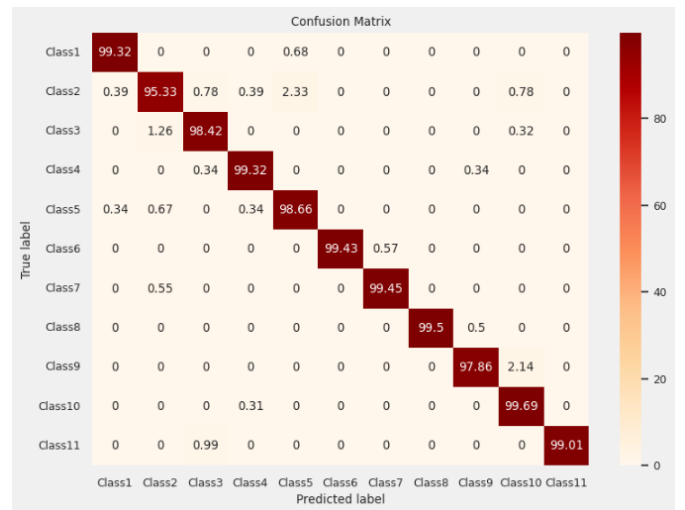


Fig. 7. Proposed model for tomato disease classification (%).

The classification report Table I provides a detailed analysis of the evaluation metrics for each tomato disease class. It includes various metrics like precision, recall, and F1-score, which assess the model’s accuracy in identifying specific diseases.

Table II provides a comprehensive analysis of the suggested model in comparison to other state-of-the-art methods tackling similar problems. The results indicate that the proposed model

TABLE I. CLASSIFICATION REPORT

Class	Precision	Recall	F1-Score	Support
Bacterial_spot	0.99	0.99	0.99	293
Early_blight	0.97	0.95	0.96	257
Late_blight	0.99	0.98	0.99	317
Leaf_Mold	0.99	0.99	0.99	296
Septoria_leaf_spot	0.97	0.99	0.98	298
Spider_mites_Two-spotted_spider_mite	1.00	0.99	1.00	174
Target_Spot	0.99	0.99	0.99	183
Tomato_Yellow_Leaf_Curl_Virus	1.00	0.99	1.00	199
Tomato_mosaic_virus	0.99	0.98	0.98	234
healthy	0.98	1.00	0.99	322
powdery_mildew	1.00	0.99	1.00	101

TABLE II. A COMPARISON OF OUR PROPOSED MODEL WITH CONTEMPORARY APPROACHES ON SIMILAR PROBLEMS

The study	Dataset	Method of Use	Accuracy
[21]	Plantvillage dataset	CNN model	95.71%
[14]	PlantVillage dataset	MobileNetV2 and NASNetMobile	97%
[13]	PlantVillage dataset	ResNet-50	96.35%
[15]	Plantvillage dataset	Wavelet-like Auto-Encoder (WAE)	98.1%
[19]	Plantvillage dataset	VGG16	95.71%
[18]	Plantvillage dataset	Inception v3	95%
[22]	Plantvillage dataset	CNN model	88.17%
This study	Plantvillage dataset	Xception and Bilinear Pooling	98.7%

achieved superior performance surpassing all other techniques mentioned in the table.

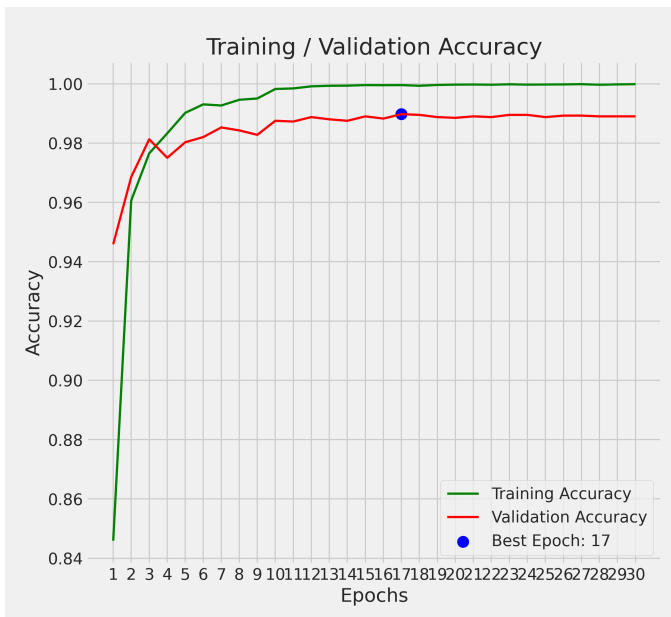


Fig. 8. Training and validation accuracy plot of the suggested model.

V. CONCLUSION

In conclusion, the proposed approach combining Xception-based CNNs and bilinear pooling demonstrates significant advancements in accurately detecting and classifying tomato diseases. With a remarkable test accuracy of 98.7%, surpassing conventional CNN approaches.

In Table I, the precision values indicate the accuracy of positive predictions for each class, ranging from 0.97 to 1.00. A higher precision value suggests a lower rate of false positive predictions.

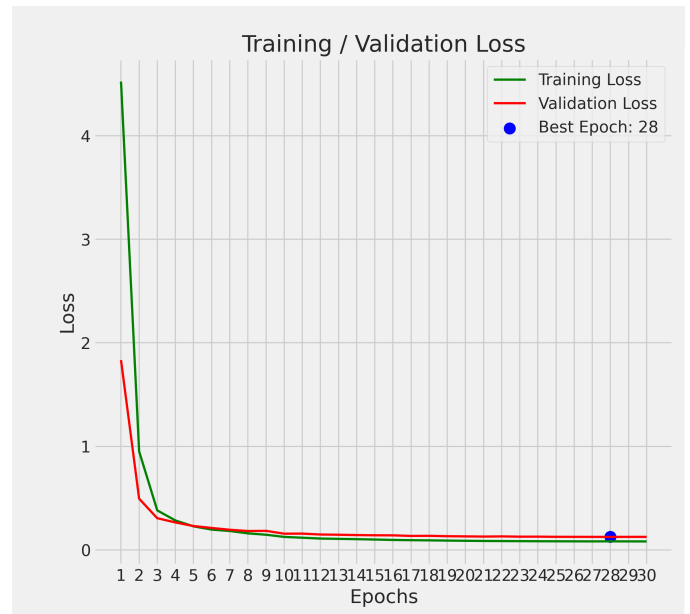


Fig. 9. Training and validation loss plot of the suggested model.

The recall values reflect the ability of the model to correctly identify positive instances for each class. The range of recall values varies from 0.95 to 1.00, indicating a high level of accuracy in capturing true positive instances.

The F1-scores, which represent the precision-recall harmonic mean, provide an overall measure of the model's performance for each class. The F1-scores range from 0.96 to 1.00, indicating a strong balance between precision and recall.

This comparison data demonstrates that the suggested model performs well across multiple classes, with consistently high precision, recall, and F1-scores. These metrics indicate the model's effectiveness in accurately identifying and classifying the different classes in the dataset.

This integrated model empowers farmers and agricultural practitioners with timely and precise disease diagnosis, enabling them to implement targeted disease management strategies and optimize yields. The successful integration of these techniques showcases the potential of advanced deep learning methods in automated tomato disease recognition, contributing to the advancement of agricultural systems.

VI. FUTURE WORKS

Building upon the innovative approach presented in this paper, future research directions in the field of agricultural disease recognition can explore several promising avenues. Firstly, investigating the generalizability of the proposed Xception and bilinear pooling model to other crops and diseases holds great potential. Additionally, refining the model's interpretability and explainability could enhance its usability by providing insights into the features and regions contributing to disease classification. Exploring techniques like attention mechanisms or saliency maps could shed light on the decision-making process of the model, enabling users to trust and fine-tune its predictions.

REFERENCES

- [1] KP Sampath Kumar, Shravan Paswan, Shweta Srivastava, et al. Tomato-a natural medicine and its health benefits. *Journal of Pharmacognosy and Phytochemistry*, 1(1):33–43, 2012.
- [2] Zhe Yan, Anne-Marie A Wolters, Jesús Navas-Castillo, and Yuling Bai. The global dimension of tomato yellow leaf curl disease: current status and breeding perspectives. *Microorganisms*, 9(4):740, 2021.
- [3] Mark Paul Selda Rivarez, Ana Vučurović, Nataša Mehle, Maja Ravnikar, and Denis Kutnjak. Global advances in tomato virome research: current status and the impact of high-throughput sequencing. *Frontiers in Microbiology*, 12:671925, 2021.
- [4] Hoseong Choi, Yeonhwa Jo, Won Kyong Cho, Jisuk Yu, Phu-Tri Tran, Lakha Salaipeth, Hae-Ryun Kwak, Hong-Soo Choi, and Kook-Hyung Kim. Identification of viruses and viroids infecting tomato and pepper plants in vietnam by metatranscriptomics. *International Journal of Molecular Sciences*, 21(20):7565, 2020.
- [5] Harikumar Pallathadka, Malik Mustafa, Domenic T Sanchez, Guna Sekhar Sajja, Sanjeev Gour, and Mohd Naved. Impact of machine learning on management, healthcare and agriculture. *Materials Today: Proceedings*, 80:2803–2806, 2023.
- [6] Aanis Ahmad, Dharmendra Saraswat, and Aly El Gamal. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology*, 3:100083, 2023.
- [7] Vipul Narayan, Pawan Kumar Mall, Ahmed Alkhayyat, Kumar Abhishek, Sanjay Kumar, Prakash Pandey, et al. Enhance-net: An approach to boost the performance of deep learning model based on real-time medical images. *Journal of Sensors*, 2023, 2023.
- [8] Bas HM Van der Velden, Hugo J Kuijf, Kenneth GA Gilhuijs, and Max A Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, 79:102470, 2022.
- [9] Jianjun Ni, Kang Shen, Yinan Chen, Weidong Cao, and Simon X Yang. An improved deep network-based scene classification method for self-driving cars. *IEEE Transactions on Instrumentation and Measurement*, 71:1–14, 2022.
- [10] Abhishek Gupta, Alagan Anpalagan, Ling Guan, and Ahmed Shaharyar Khwaja. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array*, 10:100057, 2021.
- [11] Zahid Ullah, Najah Alsubaie, Mona Jamjoom, Samah H Alajmani, and Farrukh Saleem. Effimob-net: A deep learning-based hybrid model for detection and identification of tomato diseases using leaf images. *Agriculture*, 13(3):737, 2023.
- [12] Alaa Saeed, AA Abdel-Aziz, Amr Mossad, Mahmoud A Abdelhamid, Alfadhl Y Alkhaled, and Muhammad Mayhoub. Smart detection of tomato leaf diseases using transfer learning-based convolutional neural networks. *Agriculture*, 13(1):139, 2023.
- [13] Sachin Kumar, Saurabh Pal, Vijendra Pratap Singh, and Priya Jaiswal. Performance evaluation of resnet model for classification of tomato plant disease. *Epidemiologic Methods*, 12(1):20210044, 2023.
- [14] Mehdar SAM Al-gaashani, Fengjun Shang, Mohammed SA Muthanna, Mashael Khayyat, and Ahmed A Abd El-Latif. Tomato leaf disease classification by exploiting transfer learning and feature concatenation. *IET Image Processing*, 16(3):913–925, 2022.
- [15] Mariam Moussafir, Hasna Chaibi, Rachid Saadane, Abdellah Chehri, Abdessamad El Rharras, and Gwanggil Jeon. Design of efficient techniques for tomato leaf disease detection using genetic algorithm-based and deep neural networks. *Plant and Soil*, 479(1-2):251–266, 2022.
- [16] HS Nagamani and H Sarojadevi. Tomato leaf disease detection using deep learning techniques. *International Journal of Advanced Computer Science and Applications*, 13(1), 2022.
- [17] Nishant Garg, Radhika Gupta, Maninder Kaur, Vinay Kukreja, Anuj Jain, and Raj Gaurang Tiwari. Classification of tomato diseases using hybrid model (cnn-svm). In *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pages 1–5. IEEE, 2022.
- [18] Sanjeela Sagar and Jaswinder Singh. An experimental study of tomato viral leaf diseases detection using machine learning classification techniques. *Bulletin of Electrical Engineering and Informatics*, 12(1):451–461, 2023.
- [19] Ananda S Paymode and Vandana B Malode. Transfer learning for multi-crop leaf disease image classification using convolutional neural network vgg. *Artificial Intelligence in Agriculture*, 6:23–33, 2022.
- [20] R Lohith, Kartik E Cholachudra, and Rajashekhar C Biradar. Pytorch implementation and assessment of pre-trained convolutional neural networks for tomato leaf disease classification. In *2022 IEEE Region 10 Symposium (TENSYP)*, pages 1–6. IEEE, 2022.
- [21] Marwa Abdulla and Ali Marhoon. Design a mobile application to detect tomato plant diseases based on deep learning. *Bulletin of Electrical Engineering and Informatics*, 11(5):2629–2636, 2022.
- [22] Gnanavel Sakkarvarthi, Godfrey Winster Sathianesan, Vetri Selvan Murugan, Avulapalli Jayaram Reddy, Prabhu Jayagopal, and Mahmoud Elsi. Detection and classification of tomato crop disease using convolutional neural network. *Electronics*, 11(21):3618, 2022.
- [23] Rahul Singh, Avinash Sharma, Vatsala Anand, and Rupesh Gupta. Impact of efficientnetb3 for stratification of tomato leaves disease. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, pages 1373–1378. IEEE, 2022.
- [24] Irene Sultana, Bijan Paul, Asif Mahmud, Minar Mahmud Rafi, Md Asifuzzaman Jishan, and Khan Raqib Mahmud. Automatic recognition and categorization of tomato leaf syndrome of diseases using deep learning algorithms. In *Information and Communication Technology for Competitive Strategies (ICTCS 2022) Intelligent Strategies for ICT*, pages 43–54. Springer, 2023.
- [25] Nikita Sareen, Anuradha Chug, and Amit Prakash Singh. An image based prediction system for early blight disease in tomato plants using deep learning algorithm. *Journal of Information and Optimization Sciences*, 43(4):761–779, 2022.
- [26] Emine Cengil and Ahmet Çınar. Hybrid convolutional neural network based classification of bacterial, viral, and fungal diseases on tomato leaf images. *Concurrency and Computation: Practice and Experience*, 34(4):e6617, 2022.
- [27] G Geetharamani and Arun Pandian. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering*, 76:323–338, 2019.
- [28] Zuhaib Akhtar. Xception: Deep learning with depth-wise separable convolutions. *OpenGenus IQ: Computing Expertise & Legacy*, 2022.
- [29] Long Wen, Liang Gao, Xinyu Li, and Bing Zeng. Convolutional neural network with automatic learning rate scheduler for fault classification. *IEEE Transactions on Instrumentation and Measurement*, 70:1–12, 2021.

Predicting Quality Medical Drug Data Towards Meaningful Data using Machine Learning

Suleyman Al-Showarah¹, Abubaker Al-Taie², Hamzeh Eyal Salman³, Wael Alzyadat⁴, Mohannad Alkhalaileh⁵
Software Engineering Department, Faculty of Information Technology, Mutah University, Karak, Jordan^{1,3}
Computer Science Department, Faculty of Information Technology, Mutah University, Karak, Jordan²
Software Engineering Department, Faculty of Science and IT, Al-Zaytoonah University, Amman, Jordan⁴
College of Education, Humanities and Social Sciences, Al Ain University, Al-Ain City, United Arab Emirates⁵

Abstract—This research aims to improve the process of finding alternative drugs by utilizing artificial intelligence algorithms. It is not an easy task for human beings to classify the drugs manually, as this requires much longer time and more effort than doing it using classifiers. The study focuses on predicting high-quality medical drug data by considering ingredients, dosage forms, and strengths as features. Two datasets were generated from the original drug dataset, and four machine learning classifiers were applied to these datasets: Random Forest, Support Vector Machine, Naive Bayes, and Decision Tree. The classification performance was evaluated under three different scenarios, which varied the ratio of the training and test data for both datasets, as follows: (i) 80% (training) and 20% (test dataset), (ii) 70% (training) and 30% (test dataset), and (iii) 50% (training) and 50% (test dataset). The results indicated that the Decision Tree, Naive Bayes, and Random Forest classifiers showed superior performance in terms of classification accuracy, with over 90% accuracy achieved in all scenarios. The results also showed that there was no significant difference between the results of the two datasets. The findings of this study have implications for streamlining the process of identifying alternative drugs.

Keywords—classification; alternative drugs; medical; decision tree; Support Vector Machine; Naive Bayes; Random Forest

I. INTRODUCTION

Computers have brought significant technical improvements that have resulted in the creation of huge amounts of data, particularly in pharmaceutical and healthcare systems. The availability of huge amounts of data has increased the need for data mining techniques to produce useful knowledge [1]. Accurate analyses of medical drug data are required to discover appropriate alternative medication for a patient, which is gaining with increasing data in the health care and biomedical communities [2][7].

Raw drugs data requires a clear description and interpretation for analysis purposes, this is to find the similarities between drugs that have the same properties and then to find the alternative drugs [2]. The drugs data has many different attributes collected from different sources. The heterogeneity of drug sources, and the variation of their types, made it an uneasy task for human beings to classify the drugs manually as this needs much longer time and more effort than doing it using classifiers. One of the biggest problems is that big data processing and analysis are challenging to acquire meaningful data to support an accurate medical drug practice [10]. As a result, automated medicines classifiers can assist pharmacists and clinicians in prescribing an acceptable replacement

prescription if the desired drug is unavailable, as long as the alternative drug has the same constituent name [1][29].

Quality medical drug data refers to accurate and reliable information about medications, including their uses, dosage, side effects, interactions, and other important details. This information is used by healthcare professionals, researchers, and patients to make decisions about the use and prescribing of drugs. Quality drug data is essential for ensuring safe and effective medication use and is typically obtained from reputable sources such as the FDA, the WHO, and medical literature [1].

Artificial intelligence (AI) can play a role in the collection, analysis, and dissemination of quality medical drug data. For example, AI algorithms can be exploited to mine large amounts of data from various sources, such as clinical trials, and electronic health records, to identify patterns and generate new insights about drugs [14]. AI can also be used to support drug discovery and development by identifying potential new drug candidates, predicting their efficacy and safety, and optimizing their formulation and delivery. AI models can also be used in drug safety monitoring by analyzing electronic health records, clinical trial data, and spontaneous reports to detect safety signals and identify potential risks associated with drugs [24]. Moreover, AI-based chatbots can be used to provide patients with personalized information about their medications, including dosage instructions, potential side effects, and interactions with other drugs. Overall, AI can help to improve the quality of medical drug data by enabling faster, more accurate, and more complete data analysis, and by providing new ways to access and use this information. Many ongoing AI research projects aim to improve the quality of medical drug data, such as Insilico [28], Medicine [27], Exscientia [25], DeepChem [17], and Numerate [26] all of these focused on drug discovery and development.

Pharmacies and other medical professionals can benefit from using an alternative drugs model by assisting them in classifying drugs based on their chemical properties. Such a proposed model would remarkably shorten the time to identify alternative medicines if the original is not found [1]. It also helps people who are looking for a drug with a lower price and cannot afford to purchase the original drug. So, the proposed model can help them to identify other drugs alternatives that are more affordable to them at the same time, and such options have similar chemical properties to the original option [2]. This research aims to predict alternative medical drug using AI algorithms. The alternative drug has the same chemical

characteristics as the original drugs using several features such as Ingredients, Dose Forms, and Strengths. This goal can be accomplished by proposing a conceptual approach for classifying structured medical data using Machine Learning (ML) algorithms: Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM). This aim can be accomplished by considering the following objectives: (i) to investigate the performance of the different classifiers mentioned above. (ii) to investigate the quality of raw data in terms of data amount, diversity, and minimization. (iii) to propose a conceptual approach for classifying structured medical data using machine learning.

The main contribution that can be provided by the proposed method is in the following: (i) propose an efficient method that is used to deal with semi-structured data and transform it into a structured format, which is a meaningful format used with the classifiers. (ii) provide a new way that is used to label the input drugs according to their properties based on two rules. Each labeling rule takes into account a specific combination of attributes. (iii) investigate the use of different classifiers, study their effect on the accuracy of predicting the correct drugs as well as compare their performances.

The remainder of this paper is structured as follows: Section II presents different machine learning algorithms and discusses their importance in such a study. The third Section III illustrates the related previous work. In Section IV, the proposed approach, four main steps used to build the prediction model is presented and explained. The experimental results and discussion is discussed in Section V. Lastly, Section VII presents our work's conclusion and future work.

II. BACKGROUND

Predictions from AI refer to forecasts or estimates made by an artificial intelligence (AI) system. These predictions can be made using a variety of techniques, such as machine learning, deep learning, or natural language processing [7]. The accuracy of these predictions will depend on the quality of the data used to train the AI model, as well as the complexity of the model itself. Some examples of predictions made by AI include stock market forecasting, weather forecasting, image or speech recognition, and many more [11][16]. Furthermore, the quality of the data that has been processed and analyzed by a machine learning model to make predictions or forecasts. The quality of the predictions will depend on the accuracy of the model, as well as the quality of the input data used to train the model [24].

Predicting the efficacy and safety of a medical drug is an important task in the drug development process. Artificial intelligence can be used to help predict the effectiveness of a drug by analyzing large amounts of data from preclinical and clinical trials. This includes things such as genetic data, demographics of the patient, and laboratory results [14]. In addition, artificial intelligence can be used to identify potential side effects and interactions with other drugs. One of the popular methods is using machine learning (ML) models to analyze data from drug trials and electronic health records (EHRs) to identify patterns that may indicate a drug's efficacy or potential side effects. Another approach is to use deep learning models to analyze the chemical structure of the drug and predict its

potential interactions with other molecules in the body. It is worth noting that AI-based predictions for medical drugs are still in the early stage, and many pharmaceutical companies are actively researching and developing new methods for using AI in drug development. Under the umbrella of AI in drug scope encompass, two branches of drugs are drug discovery and predicted drug.

Drug discovery is a process for identifying and developing new medications while predicting drug efficacy and safety refers to using artificial intelligence techniques to analyze data to make predictions about how a drug will perform in preclinical and clinical trials [29]. Drug discovery involves identifying potential drug targets, synthesizing and testing new compounds, and conducting preclinical and clinical trials to determine a drug's efficacy and safety. This process can take many years and involve a significant investment of time and resources [5]. On the other hand, using AI to predict drug efficacy and safety involves analyzing data from various sources, such as preclinical trial results, electronic health records, and genetic data, to identify patterns that may indicate a drug's effectiveness or potential side effects. This can be done more quickly and efficiently than traditional methods and can help reduce the cost and time associated with drug development [1] [2]. In summary, drug discovery is the process of identifying and developing new drugs from scratch, while using AI to predict drug efficacy and safety is a way to analyze existing data and make predictions about how a drug will perform in preclinical and clinical trials.

Also, this section provides the necessary background to understand how the following classifiers work: Random Forest (RF), Naive Bayes (NB), Decision Tree (DT), and Support Vector Machine (SVM).

Random Forest. Random forest is a learning algorithm for classification and regression tasks. It works by building multiple decision trees at training time, which is the cornerstone for the classification or discrimination regression processes. These multiple decision trees use RF to ensure accurate and reliable prediction [21].

Naive Bayes. Naive Bayes is one of the most famous machine learning algorithms, data analysis, and classification. Specifically, it can be characterized by rapid processing and efficiency in forecasting processes. This classifier is based on the statistical concept, Bayes' theorem. It computes the probability of a given result by verifying what is available and known as Naive because it adheres to the independence assumptions principle. As a result, the relationships between all attributes and features are thought to be independent of one another [18]. So that the Naive Bayes model is trained with the data and its characteristics available in the databases. The model then determines the type of new records and classifies them based on the data and statistics available to it. The formula for Naive Bayes is [18]:

$$P(C|X) = \frac{(P(X|C)P(C))}{(P(X))} \quad (1)$$

Decision Tree. A decision tree is a supervised learning algorithm that continuously divides data according to a specific parameter. As it is a tree that looks like a flow chart that

contains a node called the root and has no edges, while all other nodes have edges and are called leaves (also known as decision nodes) [22][3].

Support Vector Machine. Finding a hyperplane that categorizes the data points in N-dimensional space (N: the number of features) is the goal of the support vector machine algorithm (SVM) [9]. After constructing the hyperplanes, the SVM determines the boundaries between the input classes and the input elements [6].

III. RELATED WORK

There is ongoing research in using AI to predict drug efficacy and safety. Here are the most recent and relevant related works in this field.

One approach is using machine learning (ML) models to analyze data from drug trials and electronic health records (EHR) to identify patterns that may indicate the efficacy or potential side effects of a drug. Nature Medicine used an ML model to analyze data from a clinical trial of a drug to treat Alzheimer's disease and was able to predict which patients would respond well to the drug with high precision [5][20].

Another approach is to use deep learning models to analyze the chemical structure of a drug and predict its potential interactions with other molecules in the body [31]. These models can be trained on large datasets of chemical compounds and their known interactions with proteins, enzymes, and other molecules, to predict potential interactions of new compounds. A deep learning model called a graph convolutional neural network (GCNN) was trained on a dataset of known drug-protein interactions and then used to predict potential interactions of new compounds with a protein called cytochrome P450 3A4 (CYP3A4). The results showed that the model was able to predict potential interactions with high accuracy [19]. Similar research purposes are in [32], a deep learning model called a graph attention network (GAT) was trained on a dataset of known drug-protein interactions and then used to predict potential drug-protein interactions. The results showed that the model was able to predict potential interactions with high accuracy and outperformed traditional machine learning methods.

In [2], Alzyadat et al. proposed an approach to predict a targeted drug for the variety of large data structures measured by a stability scale in the preprocessing phase. Their approach performs quality data analysis using correlation methods to identify feature choices related to mapping data, which concerns the basic methods for predicting data based on the K-mean cluster and decision tree. The result of the prediction of the target drug was used as a principal component analysis (PCA) by distance value.

In [1], Al-Hgaish et al. proposed an approach based on the K-Mean algorithm to maintain the quality of medical drug data toward meaningful data in the data lake by clustering big data scope. The K-Mean clustering is used to form different clusters. Each cluster that was produced represents an alternative drug that is compatible with data lake components. The results show that the approach presented in their paper has achieved 92.7% accuracy.

In [12], Huang et al. proposed an approach to classify unknown drugs and provide assistance for drug screening during the development process. They collected a drug dataset using a web crawler. Based on this dataset, the authors derived an equation to calculate the similarity between drugs and defined similarity calculation equation parameters from a subset of the data. Drug data was categorized using the KNN (K closest neighbor) classifier based on drug similarities. The findings demonstrated that the suggested drug classification model can achieve a 77.7% accuracy value.

In [13], using the DrugBank dataset, Ibrahim et al. suggested a similarity-based machine learning system named "SMDIP". To describe the sparse feature space, they computed drug-drug similarities using an evaluation metric for the available biological and structural information on DrugBank. The chosen DDI (Drug-Drug Interaction) key features are subjected to the deployment of six different ML model types. With the following results: Precision 82%, Recall 62%, F-measure 78%, and Accuracy 79%, SMDIP has demonstrated favorable prediction performance when compared to relevant studies.

In [8], Dang et al. used data consisting of approved drugs of histamine antagonists that are connected to 26,344 Drug-Drug Interactions (DDI) pairs from the DrugBank database. Several classifiers such as Random Forest, Naive Bayes, Logistic Regression, Decision Tree, and XGBoost were used with five-fold cross-validation to approach a large-scale DDIs prediction among histamine antagonist drugs. According to the prediction performance, their model performed better than previously published works on DDI prediction with the best Precision of 78.8%, Recall of 92.1%, and F1-score of 83.8% among 19 given DDIs types.

The differences between our work with other studies are as follows:

- 1- A new methodology not used before in the previous studies to achieve the aim of this study by using the two datasets produced from the original dataset.
- 2- To the best of our knowledge, we have not come across any study conducted by using the four classifiers that are used in this study, especially, on this dataset (i.e. FDA) in order to achieve the aim.

IV. THE PROPOSED APPROACH: BUILDING PREDICTION MODEL

The proposed approach is presented in this section. In the beginning, we give a holistic view of the approach. The approach's steps are then detailed in subsequent subsections.

A. Overview

The essential steps of the proposed medicine categorization approach are explained in this section. Fig. 1 shows the four steps for the suggested model after importing the dataset to obtain an alternative drug. These steps are as follows: 1) pre-processing, 2) feature extraction, 3) applying heuristic-based rules, and 4) applying different classifiers with different scenarios. The following is an explanation of the proposed approach, which consists of the following steps:

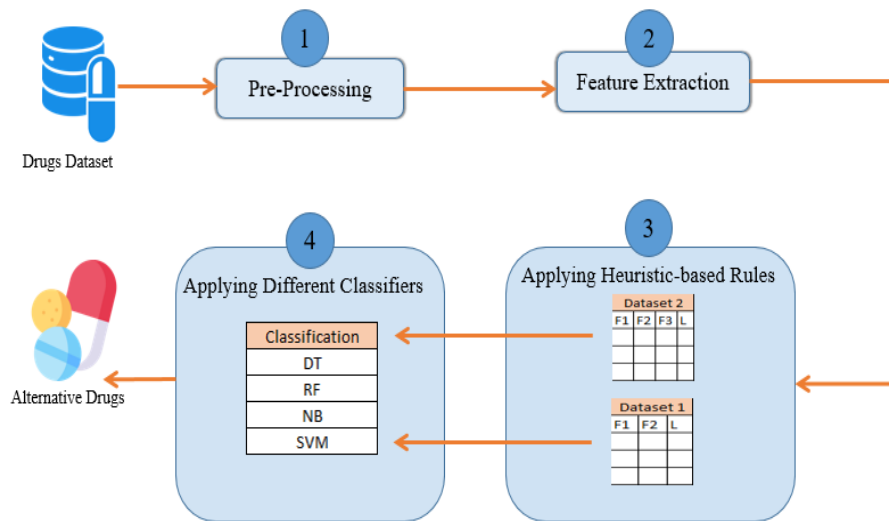


Fig. 1. An overview of the proposed approach (F: Feature, L: Label).

TABLE I. SELECTED FEATURES FOR TRAINING

Features	Type	Description
Ingredients	textual	The active ingredients of the medicine
Strengths	numerical	Active component effect
Dose Forms	numerical	The medicine dose route or form

B. Step 1: Pre-processing

This is the first step in the proposed approach, as shown in Fig. 1. Pre-processing is the step-in machine learning and natural language processing where raw data is cleaned, transformed, and organized in a format that is suitable for the model to train. This can include tasks such as tokenization, stemming, and removing stop words, as well as more complex tasks such as creating new features or dealing with missing data. Pre-processing is a crucial step in building a successful machine learning model, as the quality and structure of the input data can have a big impact on the performance of the model. In this step, you consider the content of the data (indexing) through two aspects. The first aspect performs a vertical process that indexes each attribute with missing data, incomplete data, outliers, or empty data (anything that is zero or none). The second aspect of data pre-processing is horizontal, where duplication of all rows is considered. The importance of this step is to avoid ambiguity and to increase the meaningful data set. On the other hand, the relationship between these attributes, which are best derived by feature selection will be clear [2].

C. Step 2: Features Selection

The important attributes were extracted in this step based on the previous studies and opinions of specialists of pharmacists and doctors. The database features that were picked for the investigation are displayed in Table I.

D. Step 3: Applying Heuristic-based Rules

Heuristic-based rules are a type of rule-based method used in artificial intelligence and natural language processing. These rules are based on heuristics, which are general problem-solving strategies or “rules of thumb” that are used to make decisions or solve problems. Heuristic-based rules use these rules of thumb to make decisions about how to process or understand natural language input. Heuristic-based rules are generally simple and easy to understand, but they can be prone to errors and biases. They are useful in situations where the data is well understood, and the rules can be defined to cover most cases. However, they may not be able to handle more complex or ambiguous input. An example of using heuristic-based rules in medical drug data would be to identify drug interactions based on a set of predefined rules. For example, a rule may state that if a patient is taking drug A and drug B, there is a high risk of interaction and the dosage of one or both drugs should be adjusted. Another example is extracting information from electronic medical records (EMR) using heuristic-based rules. The rules can be defined to identify specific patterns in the text, such as identifying the name of a drug, the dosage, the frequency of administration, and the duration of treatment. Once these patterns are identified, the information can be extracted and organized into a structured format for analysis. Heuristic-based rules are also used to classify clinical notes from EMR, for example, a rule can be defined that states if a patient is complaining of chest pain and shortness of breath, then it is likely a case of Angina [15]. In summary, heuristic-based rules are an efficient and cost-effective way to extract and analyze medical drug data. They are particularly useful when the data is well-defined, and the rules can be easily formulated to cover most cases. However, they may not be as robust as other methods in handling complex or ambiguous input.

In this study, two heuristic-based rules (rule 1 and rule 2) were used in the proposed approach to produce two datasets, as follows.

1) *Rule 1: Two Similar Features:* Dataset 1 was produced from the original dataset after applying rule 1, which is explained in (Equation 2 below). The content of this rule suggests that all drugs having similar ingredients, and strength values would have the same label number.

$$L1 = sim(D_i(x_1, x_2), D_j(x_1, x_2)) \quad (2)$$

From Equation 2, suppose that there are two drugs, D_i , D_j each drug has two attributes, i.e. the ingredient, and strength represented by the variables x_1 , x_2 . As rule 1 suggests that if these two drugs have the same or similar values of x_1 , x_2 , it can be said that both drugs have the same category or label (L).

2) *Rule 2: Three Similar Features:* Dataset 2 was produced using the same idea that was used to produce dataset 1. However, the content of this rule suggests that all drugs having similar ingredient, strength, and dose values would have the same label number, as stated in Equation 3.

$$L2 = sim(D_i(x_1, x_2, x_3), D_j(x_1, x_2, x_3)) \quad (3)$$

The equation can be read as follows. x_1 , x_2 , x_3 are ingredient, dose form, and strength, respectively. D_i , D_j are two drug items. Sim is similarity measurement, and L represents the label ID.

E. Step 4: Applying Different Classifiers with Different Scenarios

In this final step, the following classifiers are applied in this study: Decision Tree, Random Forest, Support Vector Machine, and Naive Bayes. These classifiers are applied individually to the datasets, which are obtained in the previous step, according to their description in the background section. The application of these classifiers on each dataset (dataset 1 and dataset 2) is done according to different scenarios. In each scenario, a percentage of the dataset's records is considered to train the prediction model while other records are used to test the prediction model.

V. EXPERIMENTS AND EVALUATION

In this section, the dataset used to evaluate the proposed approach is described, and the evaluation procedure and metrics are listed.

A. Dataset Description

The dataset utilized in this investigation is from the Food and Drug Administration (FDA) and may be found in the following link (<https://www.fda.gov/drugs/drug-approvals-and-databases/drugsfda-data-files>). This dataset consists of 14 features and 37,071 records. The selected features and the description for each feature of the dataset are based on the specialists' viewpoints as shown in Table I. The dataset was approved by FDA and this took several years and involved multiple stages, including preclinical testing, phases of clinical trials, and a review of the drug's safety and efficacy by an advisory committee [1].

The reliability of a dataset refers to the consistency and accuracy of the data it contains. In the case of datasets such as *clinicaltrials.gov* and the FDA's drug approval dataset, the

data is typically considered to be reliable as it is collected and compiled by reputable government agencies with strict oversight and regulations in place. The features and properties of drug approval datasets such as *clinicaltrials.gov* and the FDA's drug approval dataset include [1][2]:

- 1- Drug Information: these datasets contain information on drugs that are currently in development, have been approved, or have been withdrawn from the approval process. This information includes the drug's name, active ingredients, intended use, and the company developing the drug.
- 2- Study Information: these datasets also contain information on the clinical trials that have been conducted to evaluate the safety and efficacy of the drugs. This includes the study design, the number of participants, the inclusion and exclusion criteria, and the primary and secondary outcome measures.
- 3- Status Information: these datasets provide information on the current status of the drug's development and approval. This includes whether the drug is currently in preclinical testing, phase 1, 2, or 3 clinical trials, or has been approved or withdrawn by the FDA.
- 4- Search and Filter Capabilities: these datasets are searchable and can be filtered by various criteria such as drug name, condition, company, and study status.
- 5- Publicly Available: these databases are publicly available and can be accessed by anyone with an internet connection.
- 6- Regularly Updated: the data in these datasets is updated regularly as new information becomes available.

B. Research Questions and Evaluation Metrics

In this study, two research questions are answered. These questions are in the following:

- **RQ1:** To what extent the proposed approach is accurate to suggest alternative drugs? This research question aims to show the ability of the proposed approach to suggest the most suitable alternative drugs.
- **RQ2:** To what extent the proposed approach is comparable to the most recent works in the subject? This research question aims to measure the efficiency of the proposed approach when it is compared to the research works in the literature.

To address the first research question (RQ1), The obtained results are evaluated using well-known measures in this subject [4][23]. These measures are as follows: Precision, Recall, F-measure, and Accuracy. The values of these measures take a range of [0-1]. We looking to have values near to the one in all considered measures. The equations of these measures are as follows:

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

TABLE II. CLASSIFIERS RESULTS OF SCENARIO 1.

Classifier	Evaluation Measure	Dataset 1	Dataset 2
RF	F1-measure	92.34%	89.73%
	Recall	93.61%	91.62%
	Precision	91.79%	88.64%
	Accuracy	93.61%	90.27%
NB	F1-measure	97.10%	98.48%
	Recall	96.88%	98.73%
	Precision	97.59%	98.36%
	Accuracy	96.88%	98.58%
DT	F1-measure	97.95%	98.47%
	Recall	98.27%	98.73%
	Precision	97.85%	98.34%
	Accuracy	98.27%	98.73%
SVM	F1-measure	2.85%	3.44%
	Recall	9.53%	10.48%
	Precision	1.16%	2.28%
	Accuracy	8.98%	10.48%

TABLE III. CLASSIFIERS RESULTS OF SCENARIO 2.

Classifier	Evaluation Measure	Dataset 1	Dataset 2
RF	F1-measure	91.28%	87.58%
	Recall	92.83%	90.19%
	Precision	90.65%	86.09%
	Accuracy	92.83%	90.19%
NB	F1-measure	95.04%	96.94%
	Recall	95.83%	97.52%
	Precision	94.92%	96.64%
	Accuracy	95.83%	97.52%
DT	F1-measure	96.47%	96.88%
	Recall	97.10%	97.52%
	Precision	96.25%	96.54%
	Accuracy	97.10%	97.52%
SVM	F1-measure	1.45%	2.44%
	Recall	8.56%	9.24%
	Precision	0.16%	1.75%
	Accuracy	7.97%	9.49%

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (7)$$

where TP, TN, FP, and FN indicate to true positive class, true negative class, false positive, and false negative class, respectively.

The classifiers used in this study(i.e. Random Forest (RF), Decision Tree (DT), Naive Bayesian (NV), and the Support Vector Machine (SVM)) were applied to each dataset separately; dataset 1 and dataset 2. However, in the training and test dataset of each classifier, three scenarios were used in the study, as follows:

- 1- **Scenario 1:** 80% of the dataset's records are used for training while 20% of dataset's records are used for testing.
- 2- **Scenario 2:** 70% of the dataset's records are used for training while 30% of dataset's records are used for testing.
- 3- **Scenario 3:** 50% of the dataset's records are used for training while 50% of dataset's records are used for testing.

To address the second research question (RQ2), a comparison was made between the proposed approach in this study with other recent and relevant works in terms of Precision, Recall, F-measure, and Accuracy metrics.

VI. RESULTS AND DISCUSSION

This section explains the experimental results from testing the model in different scenarios. As this study was designed for three scenarios, each has two datasets based on the two rules. The results of classifiers are represented by several evaluation criteria: classification Accuracy, F-measure, Recall, and Precision.

A. Research Question (RQ1)

1) *The Results of Scenario 1:* this subsection presents the results of Scenario 1. This scenario was built based on the percentage of the samples of the dataset used in the experiment. 80% of the records were used to train the dataset while 20% of the records were used to test the dataset. This scenario applied to two datasets; the four classifiers were conducted on each dataset. Table II shows the results of Scenario 1. The results show that DT has achieved high classification accuracy for dataset 1 (98.27%), and dataset 2 (98.73%) compared to other classifiers used in the study. The high accuracy of the Decision Tree over the other classifiers can be attributed to the fact that these kinds of algorithms, i.e., the tree-based classifiers, are less prone to overfitting. At the same time, training makes them robust and rigid against outliers and misclassification [30]. Both the RF and the DT outperformed the SVM, as can be seen. Also, we can see that there was no big difference between the results of dataset 1 and dataset 2. This demonstrates that using two attributes can achieve the study's goal of finding an alternative drug with the strongest effect by using two attributes.

2) *The Results of Scenario 2:* this subsection presents the results of Scenario 2. This scenario was built based on the percentage of the samples of the dataset used in the experiment. 70% of the records were used to train the dataset, while 30% of the records were used to test the dataset. As can be seen from Table III, it can be derived the same observations noticed for scenario 1. The average performance of each one of the NB, DT, and RF was also higher than 90% which indicates that these three classifiers are good enough to handle the input data. As an observation, the SVM classifier still has minor performance outcomes compared to the other classifiers involved in the study. This is for the same interpretation mentioned in Scenario 1. Also, we can see that there was no big difference between the results of dataset 1 and dataset 2. This proves that using two attributes can achieve the aim of this study and find an alternative drug for the strongest effect of choosing the two attributes.

3) *The Results of Scenario 3:* this subsection presents the results of Scenario 3, As can be seen from Table IV. This

TABLE IV. CLASSIFIERS RESULTS OF SCENARIO 3.

Classifier	Evaluation Measure	Dataset 1	Dataset 2
RF	F1-measure	89.31%	85.60%
	Recall	91.46%	88.77%
	Precision	88.27%	83.84%
	Accuracy	91.46%	88.77%
NB	F-measure	91.64%	92.81%
	Recall	93.26%	94.41%
	Precision	91.00%	91.92%
	Accuracy	93.26%	94.41%
DT	F1-measure	93.29%	92.78%
	Recall	94.77%	94.41%
	Precision	92.54%	91.87%
	Accuracy	94.77%	94.41%
SVM	F1-measure	6.15%	5.59%
	Recall	7.18%	6.99%
	Precision	5.57%	4.86%
	Accuracy	7.18%	6.99%

scenario was built based on the percentage of the records of the dataset used in the experiment. 50% of the records were used to train the dataset while 50% of the records were used to test the dataset. Not far from the results in both Scenario 1 and 2, it can be realized that the experimental results in Scenario 3 for the tested classifiers have almost similar experimental results for the same reasons mentioned in both Scenario 1 and Scenario 2. Also, we can see that there was no big difference between the results of dataset 1 and dataset 2. This proves that using two attributes can achieve the aim of this study and find an alternative drug for the strongest effect of choosing the two attributes.

As a summary, we can note that the DT classifier outperforms other studied classifiers (RF, NV, and SVM) in terms of Precision, Recall, F-measure, and Accuracy. Also, it is noted that SVM always produced unsatisfactory results. This is based on the experimental results shown in Table II, III, and IV.

B. Research Question 2 (RQ2)

The most recent and relevant work in the literature is the work proposed by Dang et al.[8]. They conducted their study on drug data. Also, various classification algorithms were applied in their studies such as Naive Bayes, Decision Tree, Random Forest, Logistic Regression, and XGBoost. They used the Precision, Recall, F-measure to evaluate the obtained results, but we used the Accuracy measure in addition to Precision, Recall, and F-measure to evaluate our proposed model. Al-Hgaish et al.[1] conducted their study on the same as our dataset but using clustering algorithms. Their study is to maintain the quality of medical drug data toward meaningful data in the data lake by clustering big-data scope using K-Mean Algorithm. They were focused only on analyzing the dataset. Huange et al.[12] conducted their study to classify unknown drugs and provide assistance for drug screening during the development process. They collected the drug dataset using a Web crawler and applied the accuracy as a metric using the k-nearest neighbor classifier. In our study, four classifiers were used (Random Forest, Naive Bayes, Decision Tree, and Support Vector Machine). Also, two datasets were analyzed based on applying two heuristic-based rules. In addition,

three different scenarios were applied for each dataset using the following metrics in the analysis: Precision, Recall, F1-measure, and Accuracy.

Table V shows the comparison results among the most recent and relevant works (mentioned above) in this subject. It has been noted that from Table V that the proposed model has outperformed three modern methods published in the past few years in terms of accuracy Dang et al.[8], Al-Hgaish et al.[1], and Huange et al.[12]. This is because of the use of a new methodology as well as applying classifiers that have not been used before in previous studies of the dataset (i.e., the Food and Drug Administration).

VII. CONCLUSIONS AND FUTURE WORK

The goal of this study is to predict quality medical drug data toward meaningful data from an input drug dataset. The alternative drug has the same chemical characteristics as the original drugs have several features: ingredients, dose forms, and strengths. This aim can be accomplished by considering the following objectives: (i) to investigate the performance of different classifiers (i.e., Decision Tree, Random Forest, Support Vector Machine, and the Naive Bayesian) on the drugs dataset. (ii) to investigate the quality of raw data in terms of data amount, diversity, and minimization. (iii) to propose a conceptual approach for classifying structured medical data using machine learning. The experiments were conducted on three scenarios for the following classifiers: Decision Trees, Random Forest, Support Vector Machine, and Naive Bayesian. The obtained results indicated that the Decision Tree, Naive Bayes, and Random Forest classifiers showed superior performance in terms of classification accuracy, with over 90% accuracy achieved in all scenarios. The results also showed that there was no significant difference between the results of the two generated datasets. The findings of this study have implications for streamlining the process of identifying alternative drugs. When it comes to the performance of the Support Vector Machine, it can be realized that it has a major degradation in performance.

Future work involves exploring the use of more advanced machine-learning techniques to improve the accuracy and performance of the classifiers. Another avenue for further research would be to include more features and variables in the analysis to provide a more comprehensive evaluation of the drugs. Additionally, it would be beneficial to compare the results of this study with other existing drug classification systems to identify any areas for improvement. Finally, conducting user studies and gathering feedback from medical professionals could provide valuable insights into the real-world applicability of the proposed approach and identify any potential limitations.

REFERENCES

- [1] A. Al-Hgaish, W. Alzyadat, M. Al-Fayoumi, A. Alhroob, and A. Thunibat. Preserve quality medical drug data toward meaningful data lake by cluster. *International Journal of Recent Technology and Engineering*, 8:270–277, 2019.
- [2] W. Alzyadat, M. Muhairat, A. Alhroob, and T. Rawashdeh. A recruitment big data approach to interplay of the target drugs. *Int. J. Advance Soft Compu. Appl*, 14(1), 2022.

TABLE V. COMPARISON TABLE WITH RELATED WORKS.

	Precision	Recall	F-Measure	Accuracy
Dang et al. (2021)[8]	78.8%	92.1%	83.8%	—
Al-Hgaish et al. (2019)[12]	—	—	—	92.7%
Huang et al. (2017)[1]	—	—	—	77.7%
Our Proposal	97.85%	98.27%	97.95%	98.27%

- [3] P. Argentiero, R. Chin, and P. Beaudet. An automated approach to the design of decision tree classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(1):51–57, 1982.
- [4] R. A. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc., USA, 1999.
- [5] D. S. Battina. The role of machine learning in clinical research: Transforming the future of evidence generation. *FUTURE*, 4(12), 2017.
- [6] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, 408:189–215, 2020.
- [7] H. S. Chan, H. Shan, T. Dahoun, H. Vogel, and S. Yuan. Advancing drug discovery via artificial intelligence. *Trends in pharmacological sciences*, 40(8):592–604, 2019.
- [8] L. H. Dang, N. T. Dung, L. X. Quang, L. Q. Hung, N. H. Le, N. T. N. Le, N. T. Diem, N. T. T. Nga, S.-H. Hung, and N. Q. K. Le. Machine learning-based prediction of drug-drug interactions for histamine antagonist using hybrid chemical features. *Cells*, 10(11), 2021.
- [9] R. Gandhi. Support vector machine — introduction to machine learning algorithms. 2018.
- [10] A. Halevy, F. Korn, N. F. Noy, C. Olston, N. Polyzotis, S. Roy, and S. E. Whang. Goods: Organizing google’s datasets. SIGMOD ’16, New York, NY, USA, 2016. Association for Computing Machinery.
- [11] E. Hamadaqa, A. Abadleh, A. Mars, and W. Adi. Highly secured implantable medical devices. In *2018 International Conference on Innovations in Information Technology (IIT)*, pages 7–12, 2018.
- [12] D. G. Huang, L. Guo, H. Y. Yang, X. P. Wei, and B. Jin. Chemical medicine classification through chemical properties analysis. *IEEE Access*, 5:1618–1623, 2017.
- [13] H. Ibrahim, A. M. El Kerdawy, A. Abdo, and A. Sharaf Eldin. Similarity-based machine learning framework for predicting safety signals of adverse drug–drug interactions. *Informatics in Medicine Unlocked*, 26:100699, 2021.
- [14] K.-K. Mak and M. R. Pichika. Artificial intelligence in drug development: present status and future prospects. *Drug discovery today*, 24(3):773–780, 2019.
- [15] G. G. Marewski JN. Heuristic decision making in medicine. *Dialogues Clin Neurosci*. 2012, 14(1):77–89, 2022.
- [16] A. Mars, A. Abadleh, and W. Adi. Operator and manufacturer independent d2d private link for future 5g networks. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPs)*, pages 1–6, 2019.
- [17] A. J. Minnich, K. McLoughlin, M. Tse, J. Deng, A. Weber, N. Murad, B. D. Madej, B. Ramsundar, T. Rush, S. Calad-Thomson, et al. Ampl: a data-driven modeling pipeline for drug discovery. *Journal of chemical information and modeling*, 60(4):1955–1968, 2020.
- [18] T. R. Patil and S. S. Sherekar. Performance analysis of naive bayes and j 48 classification algorithm for data classification. 2013.
- [19] M. Qiu, X. Liang, S. Deng, Y. Li, Y. Ke, P. Wang, and H. Mei. A unified gnn model for predicting cyp450 inhibitors by using graph convolutional neural networks with attention mechanism. *Computers in Biology and Medicine*, 150:106177, 2022.
- [20] S. Qiu, M. I. Miller, P. S. Joshi, J. C. Lee, C. Xue, Y. Ni, Y. Wang, D. Anda-Duran, P. H. Hwang, J. A. Cramer, et al. Multimodal deep learning for alzheimer’s disease dementia assessment. *Nature communications*, 13(1):1–17, 2022.
- [21] M. Ristin-Kaufmann. Large-scale image recognition with random forests. ETH-Zürich, 2015.
- [22] S. Safavian and D. Landgrebe. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(3):660–674, 1991.
- [23] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., USA, 1986.
- [24] M. Schaeperl and R. A. Denny. Ai-based protein structure prediction in drug discovery: Impacts and challenges. *Journal of Chemical Information and Modeling*, 62(13):3142–3156, 2022.
- [25] E. Smalley. Ai-powered drug discovery captures pharma interest. *Nature Biotechnology*, 35(7):604–606, 2017.
- [26] N. Stephenson, E. Shane, J. Chase, J. Rowland, D. Ries, N. Justice, J. Zhang, L. Chan, and R. Cao. Survey of machine learning techniques in drug discovery. *Current drug metabolism*, 20(3):185–193, 2019.
- [27] N. J. Sucher. The application of chinese medicine to novel drug discovery. *Expert opinion on drug discovery*, 8(1):21–34, 2013.
- [28] G. C. Terstappen and A. Reggiani. In silico research in drug discovery. *Trends in pharmacological sciences*, 22(1):23–26, 2001.
- [29] J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, et al. Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6):463–477, 2019.
- [30] K. Vijayakumar and C. Saravanakumar. *Multilevel Mammogram Image Analysis for Identifying Outliers: Misclassification Using Machine Learning*, pages 161–175. Springer Singapore, Singapore, 2021.
- [31] J. You, R. D. McLeod, and P. Hu. Predicting drug-target interaction network using deep learning model. *Computational biology and chemistry*, 80:90–101, 2019.
- [32] L. Zangari, R. Interdonato, A. Calió, and A. Tagarelli. Graph convolutional and attention models for entity classification in multilayer networks. *Applied Network Science*, 6(1):1–36, 2021.

Incorporating Learned Depth Perception Into Monocular Visual Odometry to Improve Scale Recovery

Hamza Mailka, Mohamed Abouzahir, Mustapha Ramzi

High School of Technology of Salé Laboratory of Systems Analysis-Information Processing
and Industrial Management (LASTIMI),
Mohammed V University in Rabat, Morocco

Abstract—A growing interest in autonomous driving has led to a comprehensive study of visual odometry (VO). It has been well studied how VO can estimate the pose of moving objects by examining the images taken from onboard cameras. In the last decade, it has been proposed that deep learning under supervision can be employed to estimate depth maps and visual odometry (VO). In this paper, we propose a DPT (Dense Prediction Transformer)-based monocular visual odometry method for scale estimation. Scale-drift problems are common in traditional monocular systems and in recent deep learning studies. In order to recover the scale, it is imperative that depth estimation to be accurate. A framework for dense prediction challenges that bases its computation on vision transformers instead of convolutional networks is characterized as an accurate model that is utilized to estimate depth maps. Scale recovery and depth refinement are accomplished iteratively. This allows our approach to simultaneously increase the depth estimate while eradicating scale drift. The depth map estimated using the DPT model is accurate enough for the purpose of achieving the best efficiency possible on a VO benchmark, eliminating the scaling drift issue.

Keywords—Visual odometry; scale recovery; depth estimation; DPT model

I. INTRODUCTION

The greatest anticipated technological advancement in the near future is autonomous ground vehicles (AGV), which operate entirely automatically. A vehicle operating autonomously requires precise and reliable information regarding its position for a successful navigation [1], [2], [3], [4]. There are currently many popular methods of providing comparatively reliable positioning information, such as the Global Navigation Satellite System (GNSS), Visual Odometry, and the Inertial Navigation System (INS) [5], [6], [7], [8], [9], [10], [11].

A growing interest in autonomous vehicles has led to well-developed novel approaches based on VO. Different approaches have been well studied since they estimate the position and orientation of moving objects based on the analysis of image sequences [12], [13], [14], [15]. A precise VO system is one of the most crucial techniques in the area of mobile robots. [16], [17], [18], [19]. The way conventional monocular VO systems operate is by assuming that the scale is one or by using ground truth for an approximation of the scale. Due to significant drift, monocular VO systems cannot operate on image sequences without ground truth or estimate the pose with significant drift [20], [21], [22], [23].

Despite the fact that several traditional monocular VO systems have been developed, they have still performed poorly or are unable to work in some conditions, like monotonous scenes that lack visible texture information or large-scale camera movements. When it comes to learning-based VO systems, they are developed by training neural networks in supervised or unsupervised manners through end-to-end pose estimation [24], [25], [26], [27]. Moreover, the efficiency of networks completely determines how accurate pose estimation is. Even with numerous training datasets and a network structure optimized, it is unavoidable for a network to encounter issues such as insufficient accuracy when estimating rotational pose.

Over the years, a lot of work has gone into developing a reliable and precise VO system. In terms of traditional VO algorithms, two main types exist: feature-based [13], [28] and direct methods [29], [30]. Calibration of the camera, identifying and matching features, rejecting outliers (using RANSAC), estimating motion, and estimating scale are typical components of feature-based techniques (e.g., Bundle Adjustment). However, finding the right features to reconstruct certain motions is still difficult. The motion of the pixel is tracked, and pose predictions are obtained by minimizing photometric error, so it is highly sensitive to light variations. Additionally, the classic monocular VO's absolute scale estimation requires the use of additional data or knowledge (such as the camera's height). In monocular systems, obtaining scale information is complicated and typically depends on an earlier, predefined absolute reference. A reference scale can be provided by integrating with some other sensors, like an inertial measurement unit. Scale drift [31] is often addressed by local optimization techniques like bundle adjustment and loop closure detection. In addition, researchers employ the depth estimation [32] from images to approximate the scale and adjust the calculated translation in addition to other approaches, like a ground plane estimation using the camera height, which is assumed to remain stable during motion.

The vast amount of training data (ground truth) required by supervised deep learning methods is usually collected with RGB-D cameras indoors and 3D laser scanners. Nevertheless, since ground truth is required, the supervised technique has a number of drawbacks. At first, the sensors' own inaccuracy and noise may have an impact on the network. Second, these sensors cannot record high-resolution information as well as images since their measurements are often sparser. Lastly, those sensors may not be able to obtain ground truth in some

locations. Because of this, researchers have begun to pay greater attention to unsupervised approaches that only need training data. In this context, the goal of our contribution is to propose a reliable localization that just requires images from a monocular camera in order to obtain an estimation of motion. The suggested strategy deals with the problem of scale estimation by using a dense prediction transformer model to estimate the depth map of the environment. As shown in this work, according to KITTI odometry benchmarks, the system's scale estimation performs in a manner comparable to that of the state-of-the-art.

The remainder of this paper is organized as follows: Section II represents a brief summary of related work. The DPT model utilized in our paper is summarized in Section III. In Section IV, we provide details on our end-to-end approach. Section V shows experimental results on the KITTI. In Section VI, the paper's contributions are summarized, as well as it concludes with some recommendations for future work.

II. RELATED WORK

Numerous studies have investigated how depth can be estimated from images employing stereo and monocular images, or multi-view images. By using conventional and traditional techniques in a single-view image, it is difficult to recognize the structure of the scene. Luckily, since the innovative research of [33], deep learning has progressed greatly in the computer vision field. Most CNN-based depth-map prediction approaches are supervised. To learn parameters, these techniques require more than one labeled dataset. We will look at how to solve the scale estimation problem in the following parts: In this section, we provide a brief summary of the most closely related work that is needed to assess the scene depth estimation and camera motion prediction.

A. Recovering Camera Poses and Depth using Conventional Techniques

Researchers in computer vision have long been interested in recovering depth-maps and camera poses. A learning depth approach based on 2D to 3D image conversion using examples is proposed by Konrad et al. [34]. They create an efficient and simplified version of the current 2D to 3D frame conversion algorithms. A feasible depth generation method from sequences of images was described in [35] employing auxiliary data from non-parametric depth sampling. The performance of this method was superior to all current standard depth methods. Camera posture research another important topic of study in the discipline of computer vision, has a great success using conventional methods. The most well-known conventional approach that is used to estimate camera pose using images is called ORB-SLAM [36]. It uses the feature matching approach for mapping and localization combined with a single monocular image. The process of this method has four stages: loop closure, tracking, mapping, and re-localization. But each stage needs to be carefully planned. Gao et al. [37] expanded this knowledge to build reconstructions of 3D objects from 2D images based on motion techniques. A method for predicting 3D structures and camera projections was put out in [38]. The strategy is in the area of estimating 3D symmetric objects from 2D symmetric perspectives and forms using numerous intra-class objects as an input model. It was

suggested by Ma et al. [39] that in remote sensing imagery, rigid and non-rigid structures can be matched using a locally linear transformation model. To determine scene depth, all of the approaches mentioned above either rebuild 3D geometry or establish pixel-by-pixel correspondences between input views. Nevertheless, the input data for these methods is multi-view images.

B. A Supervised Learning Approaches using Monocular Images

Since we can't obtain the structural properties from a single view image, determining a depth-map using a monocular camera is a difficult issue. Depth estimation has recently been viewed by some academics as a supervised learning approach. A network with two factors was proposed by Eigen et al. [40], the first of which assesses the scene's overall structure and the second of which refines it using local information. As one of the few papers using deep CNN to estimate scene depth using monocular images. Three separate computer vision issues were handled simultaneously via a framework that Eigen et al. [41] created (prediction of surface normals, depth estimates, and semantic identification) based on prior research. A completely convolutional architecture was suggested by Laina et al. [42] to describe the uncertain mapping between depth maps and monocular pictures. Li et al. [43] introduced a multi-streamed CNN architecture for depth estimation that is quick to train. Yan et al. [44] used a reference as the surface normal to aid in the monocular depth estimation problem. Up to this point, some studies on monocular depth prediction have combined CNNs and Random Forests. In certain studies on monocular depth prediction, CNNs and random forests were merged. Regression based on deep CNN characteristics was used by Li et al. [45] to overcome this issue, together with conditional random fields for post-processing refinement. Using only one image as a source of depth prediction, Roy et al. [46] introduced a deep regression forest approach that blends CNNs with random forests. As a result of depth data being continuous, Liu's formulation of depth prediction as the "random field learning with continuous conditions" problem [47] was developed. Although the aforementioned techniques have shown precise monocular-depth prediction, the ground truth is used as a basis for training, which limits the model's capacity for generalization.

C. An Unsupervised Learning Approaches using Monocular Images

Several unsupervised learning techniques that address the monocular depth estimation issue have recently been introduced to get over the ground-truth issue. Garg et al. [48] built a CNN to approximate the complex non-linear transformation that turns stereo images into depth maps using input camera motions. The proposed method of [49], [50] constructed a model based on Garg's work to incorporate a spatial smoothness loss into the unsupervised optical flow total loss function. Their efforts and outcomes are comparable. When training, Godard et al. [51] approached the difficulty of predicting depth as an issue with image reconstruction using epipolar geometry constraints. To determine the relationship between the rectified stereo images, a loss function is developed. In a semi-supervised manner, Kuznetsov et al. [52] employed

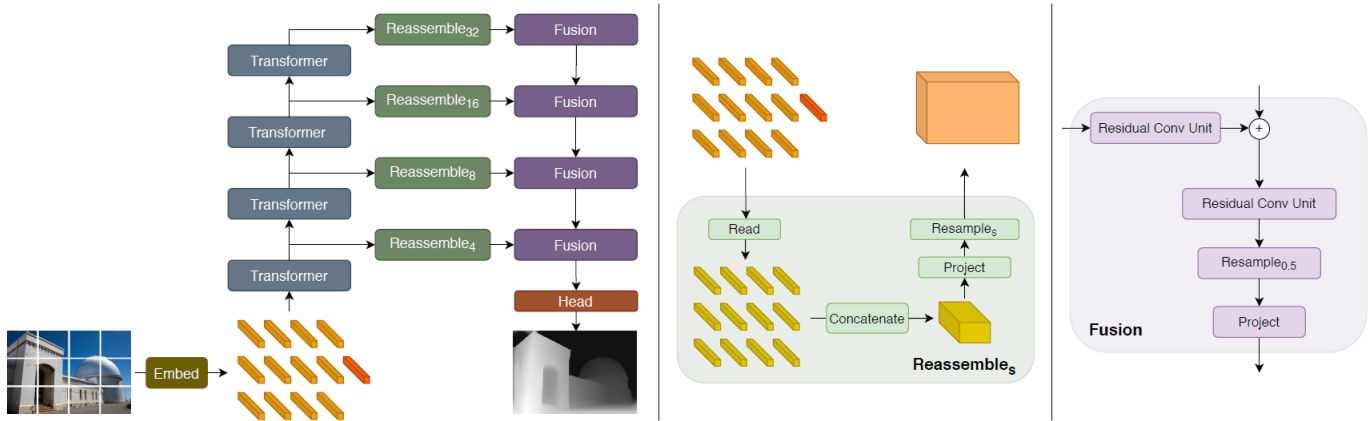


Fig. 1. The architecture of dense prediction transformer model that we used in our approach to estimate disparity maps from monocular cameras. This enables the recovery of precise metric estimates [61].

projected inverse and sparse ground truth depth data. Their models depend on a 3D laser sensor and camera that are precisely calibrated externally. A learning approach called CCRFN (Convolutional Conditional Random Field Network) was proposed by Yan Hua and Hu Tian [53] For estimating depth and identifying features by using the learning approach. It offers two benefits: first, it doesn't require manually created features, and second, it uses the relationship between individual features to estimate depth. Based on the research of Zhou[54], Yin et al. [55] introduced the GeoNet architecture for unsupervised learning, which jointly predicts optical flow, monocular depth, and dynamic object detection. Luo et al. [56] from SenseTime Research suggested stereo matching as the first sub-issue to address after the monocular depth estimation method. The above unsupervised learning approaches were trained on mono-image sequences or using stereo images with precise calibration. Temporal information cannot be fully utilized by stereo pictures. Due to depth ambiguity, which can occur in monocular images, different depths may correlate to objects that appear to be similar in the image. Although these unsupervised models succeeded in their objective of estimating scene depth without the need for ground truth, they have received little attention for their joint use of stereo and monocular sequences for depth prediction.

III. DPT MODEL

Due to the quick advancement of computer technology and digital imaging sensors. The camera sensor is progressively becoming more advantageous, and as a result, navigation using visual assistance and its related combined system have emerged as a significant component of the integrated navigation system. Since Transformer was so successful in natural language processing (NLP) [57], the computer vision community has given it a lot of attention. It has recently demonstrated exceptional performance on a variety of computer vision tasks, including semantic segmentation, object identification, imaging classification, and depth estimation. The standard architecture for dense prediction is fully-convolutional networks [58], [59]. Although many variations of this fundamental pattern have been presented throughout time, all current architectures use convolution and subsampling as their core components to learn

multi-scale models that can make use of a sufficiently wide context. When trained on enormous datasets and deployed as high-capacity architectures, transformer models have proven particularly effective. Attention processes have been adapted to image analysis in a number of publications. In particular, it has recently been shown that a direct application of effective token-based transformer designs in NLP may produce competitive performance on image categorization [60]. This work's most important finding was that, similar to transformer models in NLP, visual transformers require a substantial quantity of training data to reach their full potential.

In contrast to the state-of-the-art CNN-based method, Ranftl et al. [61] reported improved relative performance using the dense prediction transformer (DPT) model for monocular depth estimation. This is why we decided to estimate the depth map using the DPT network since precise depth estimation improves scale estimation [20]. A ViT serves as the basis of the DPT model. The frame is divided into regions, which are subsequently incorporated as flattened depictions of ResNet-50 network-derived features [61]. The CNN feature extractor's embedding step turns the model into a hybrid one (DPT-Hybrid Architecture [61]). Following the original terminology of the transformer architecture, we shall refer to the embedded patches as tokens and the image patches as "words" in NLP tasks. The tokens are transformed using layers composed of multi head self-attention blocks. A reassemble operation is used to reassemble the output of the transformer layers, and then fusion blocks are used to gradually fuse the characteristics. In the embedding stage, the DPT-Hybrid architecture makes use of features that were taken from layers of the ResNet 50, and Fig. 1 depicts the entire dense prediction transformer model.

IV. PIPELINE OF MONOCULAR VISUAL ODOMETRY

The pipeline of monocular visual odometry is based basically on the DPT model to calculate depth estimation. Feature detection using a fast feature detection approach, matching features with optical flow, depth estimation by DPT, scale, and motion estimation blocks are the primary block components of the pipeline shown schematically in Fig. 2, which is also illustrated in Algorithm 1.

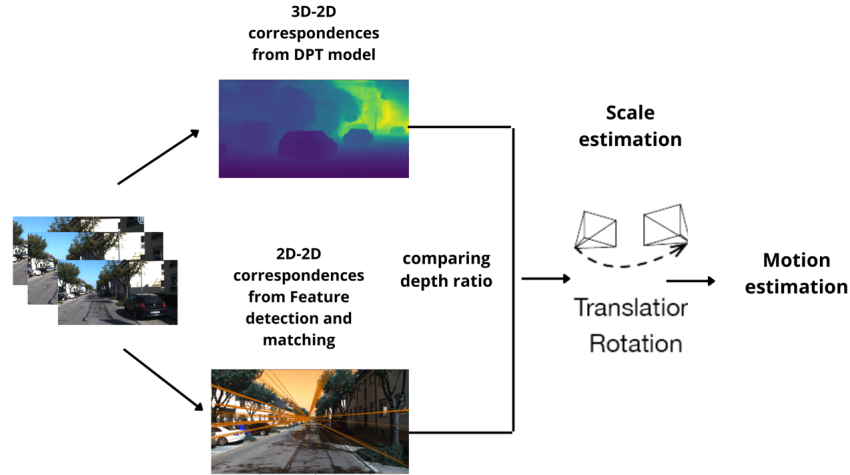


Fig. 2. The proposed approach's architecture estimates the scale from a deep-learning model to estimate the disparity map from a monocular camera. This enables the recovery of precise metric estimates.

Algorithm 1 Proposed Visual Odometry algorithm

Require: *Model* : *dpt_hybrid_kitti-cb926ef4.pt*

Frames : $[F_1, \dots, F_k]$

Ensure: Vehicle poses: $[T_1, T_2, \dots, T_k]$

- 1: **Initialization:** $n=2, N=\text{number_Of_Frame}$
 - 2: $\text{Prev_Feature}=\text{FastFeatureDetection}(F_1)$
 - 3: $\text{Last_Frame}=F_1$
 - 4: **while** $n \leq N$ **do**
 - 5: $\text{Prev_Feature}, \text{Cur_Feature}=\text{featureTracking}(\text{Last_Frame},$
 - 6: $F_n, \text{Prev_Feature})$
 - 7: Compute E using Prev_Feature and Cur_Feature
 - 8: compute $[R, t]$ using Essential matrix E
 - 9: Get Depth frame prediction D_n
 - 10: Get D'_n from Triangulation between Prev_Feature and Cur_Feature
 - 11: α : Scale estimation from comparison between D'_n, D_n
 - 12: **if** $|\alpha - \text{absoluteScale}| < \xi$ **then**
 - 13: $T_n = [R, \alpha t]$
 - 14: **else**
 - 15: 3D-2D correspondences using $D_n, \text{Prev_Feature}$ and Cur_Feature
 - 16: Compute $[R, t]$ from PnP
 - 17: **end if**
 - 18: $n++$
 - 19: $\text{Last_Frame}=F_n$
 - 20: $\text{Prev_Feature}=\text{Cur_Feature}$
 - 21: **end while**
-

A. 2D-2D Correspondences

Monocular VO uses a single camera to combine images in an effort to progressively estimate an agent's motion. Epipolar geometry is one of the fundamental approaches that can be used to compute the pose from frame sequences using monocular or stereo cameras. Epipolar geometry is based on many steps, from 2D-2D correspondence to solving the essential matrix and the fundamental matrix (E, F). From the intense optical flow, the 2D-2D correspondences are recovered. Given a pair of frames, $(F_k; F_{k+1})$, optical flow can be

used to characterize the feature variation of time and provide correspondences for all the features that were derived from F_i and their correspondences in F_j . For the purpose of solving the fundamental matrix F and the essential matrix E , epipolar constraint is used based on the intrinsic calibration matrix K also indicates that the projection characteristics of the camera, where $F_k = K^{-T} E_k K^{-1}$ and the motion of the vehicle can be estimated using the following equation:

$$T = \begin{bmatrix} R_k & t_k \\ 0 & 1 \end{bmatrix} \quad (1)$$

with $R_k \in SO(3)$ and $t_k \in \mathbb{R}^{3 \times 1}$ are the rotation matrix and translation vector, respectively, that illustrate how the camera rotated and translated from instant $k - 1$ to k . The following are the manners in which the camera motion is associated with the essential matrix:

$$E = [t]_{\times} R \quad (2)$$

1) *Fast feature detection:* In the feature detection stage, interesting features in each frame, such as corners, are found. These locations are known as keypoints or features, and the next frames should be able to clearly identify them so that feature matching may be used.

Rosten and Drummond [62] developed the FAST feature detector (Features from Accelerated Segment Test). Fast criteria for interest point identification have grown in popularity as cutting-edge techniques with strict real-time limitations. In Fig. 3, a feature is shown at pixel p if the intensities of at least nine surrounding pixels in a 16 pixel circle are all either lower than or higher than $I(p)$ by a threshold score. Training a decision tree further accelerated the algorithm, which examines candidate pixels into corners and is not based on as few pixels as possible. The procedure was accelerated even more by instructing a decision tree to look at as few pixels as possible to identify whether a candidate pixel is a corner or not. The segment test characteristics cannot be directly suppressed using

a non-maximal approach because no corner response function has been constructed. Therefore, a scoring function must be calculated for each detected corner in order to delete any corners that have an adjacent corner with a higher $C.C$ is provided by:

$$C = \max \left(\sum_{i \in S_{\text{bright}}} |I_{p \rightarrow i} - I_p| - t, \sum_{i \in S_{\text{dark}}} |I_p - I_{p \rightarrow i}| - t \right) \quad (3)$$

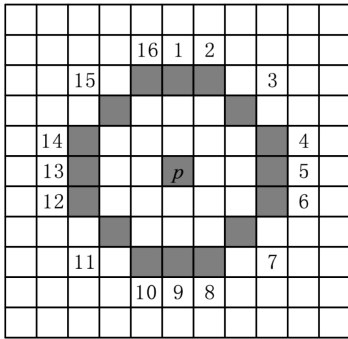


Fig. 3. Fast Feature Detection.

An analysis of similarity is performed at the end of the feature extraction process to compare each keypoint in a frame to all other keypoints in the following frame. In order to match detected features, the Lucas-Kanade method is used to calculate optical flow between frames iteratively.

2) *Optical flow*: Compared to other computer vision issues, ego-motion estimation has a fundamentally different basis, which places a greater focus on geometric motion inside individual video frames. The camera's output frame varies over time and could be seen as a function of time, and the assumption of photometric constancy is the basis for the optical flow computation. Alternatively, every frame has the same spatial location and a predetermined pixel intensity value. The following characteristics apply in the case of a pixel shifting to $(x + \Delta x, y + \Delta y)$ at a time of $t + \Delta t$:

$$M(x + \Delta x, y + \Delta y, t + \Delta t) = M(x, y, t) \quad (4)$$

On the left side of Eq. (4), we may carry out the first-order Taylor expansion:

$$M(x + \Delta x, y + \Delta y, t + \Delta t) \approx M(x, y, t) + D_x \Delta x + D_y \Delta y + D_t \Delta t \quad (5)$$

where:

$\frac{\partial M}{\partial x}$, $\frac{\partial M}{\partial y}$, and $\frac{\partial M}{\partial t}$ are the frame's gradients D_x , D_y , and D_z in the x , y , and t axes, respectively. $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$ are the pixels' rates of movement on the x and y axes, respectively. and The future grayscale equals the prior one based on photometric consistency, so:

$$uDx + vDy = -Dt \quad (6)$$

Eq. (6) can be expressed as a matrix:

$$[D_x, D_y] \begin{bmatrix} u \\ v \end{bmatrix} = -D_t \quad (7)$$

The conventional approach is to use the Lucas-Kanade (LK) method to introduce the least squares solution to establish the u, v pixel motion. By doing this, we can determine how quickly pixels change between frames. Fig. 4 shows an example of feature tracking using optical flow.

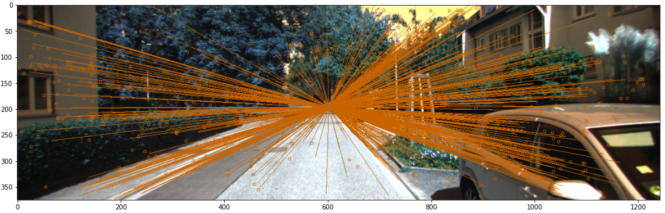


Fig. 4. Feature tracking using optical flow.

B. 3D-2D Correspondences

It is possible to construct 2D-2D and 3D-2D correspondences given a depth prediction from the DPT model and the features extracted using fast feature detection. Either PnP (3D-2D) or the essential matrix can be used to solve the relative camera pose.

1) *Depth estimation*: The problem of dense regression is frequently used to model a monocular depth estimate. Massive datasets can be formed from sources of data that already exist if certain considerations are made in how many depth representations are combined into a single representation and common ambiguities (like scale ambiguity) are addressed properly in the training loss. Since it is well known that transformers only perform to their greatest potential when a wealth of training data is provided, Our research primarily relied on monocular depth estimation using the DPT model.

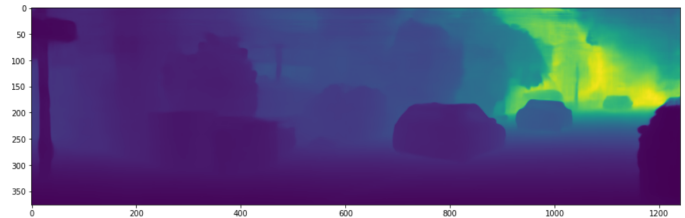


Fig. 5. The DPT model was used to create a depth view of the KITTI dataset image.

Fig. 5 shows an example of using the DPT model to create a depth view based on the Kitti dataset. A convolutional decoder is used by DPT to gradually merge tokens from various stages of the vision transformer into full-resolution predictions. The translation vector can be corrected and the relative scale in MVO estimated in a variety of ways. The scale can be estimated using the depth information and also using the earlier knowledge of camera height. The scale recovery approach used by Zhan et al [63]. is based on CNN depth estimation for lining

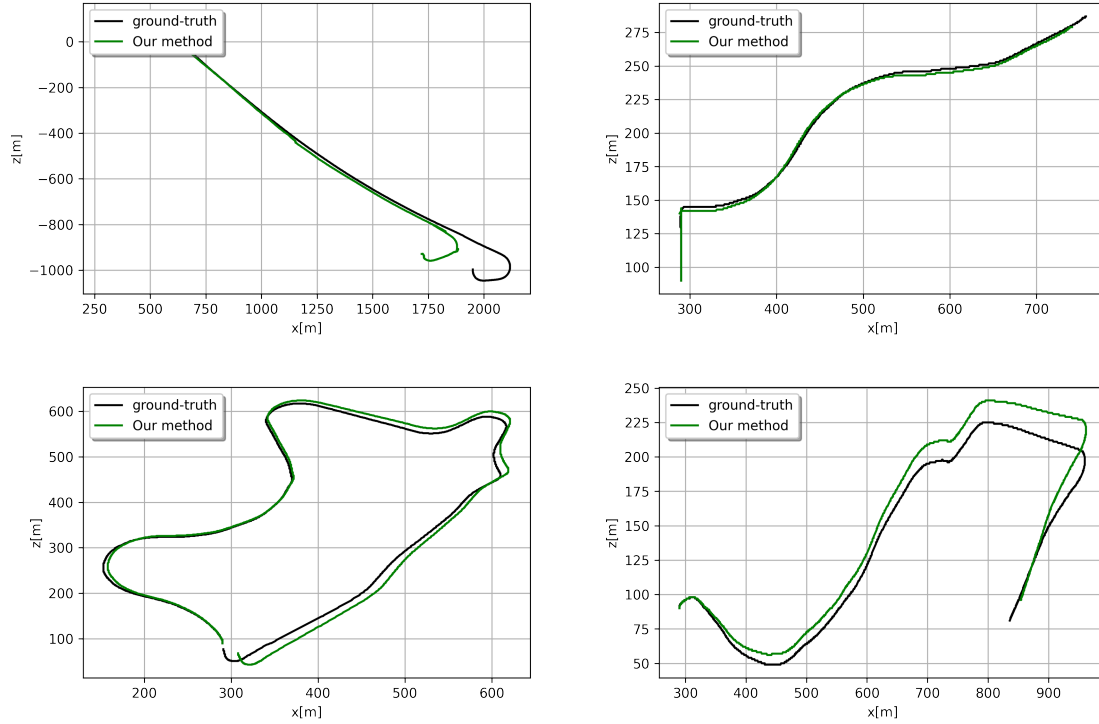


Fig. 6. Localization results of depth visual odometry compared to ground-Truth of KITTI dataset.

up depth based on a triangulated approach with the estimated depth map from deep learning. The realistic 3D structures are presumed to be known in this scenario since the depth sensor is supposed to be the deep learning model. Once the triangulation procedure has removed any earlier outliers, let M represent the number of keypoints that are still matching. A vector representing depth ratios is utilized to establish the scale, as is a RANSAC regressor.

$$\mathbf{D} = \left[\frac{\hat{D}_0}{D_0}, \frac{\hat{D}_1}{D_1}, \dots, \frac{\hat{D}_M}{D_M} \right]^T \quad (8)$$

2) *PnP(Perspective from n-point)-based motion estimation:* To address this pose estimation problem using 3D-to-2D correspondences, either a nonlinear strategy (perspective from n points, PnP approach) or a linear technique (DLT-based estimation) can be used. The keypoints' depth is considered to be known given the estimated depth map produced by the deep learning model. Therefore, using a 2D projection of the corresponding 3D point, the PnP predicts the camera motion \mathbf{T}_k as follows:

$$\arg \min_{\mathbf{T}_k} \sum_i \left\| \mathbf{K} (\mathbf{R} \mathbf{P}_{k-1}^i + \mathbf{t}) - \mathbf{x}_k^i \right\|_2 \quad (9)$$

V. RESULTS AND DISCUSSION

As a primary method for evaluating our trajectories, we employ the KITTI relative error metric. According to the error metrics, we compute the average RMSE error for the rotational

r_{error} and translational t_{error} errors using different sequences of KITTI Dataset. The relative pose error, which is particularly helpful for the evaluation of visual odometry approaches since it correlates to the drift of the trajectory, assesses the local accuracy of the trajectory over a set time period Δ . at time step i let's define the relative pose error matrix as follows:

$$E_i := \left(\hat{P}_i^{-1} \hat{P}_j \right)^{-1} \left(P_i^{-1} P_j \right) \quad (10)$$

The relative pose error matrix m is obtained from a sequence of N camera poses where $M = N - \Delta$ where the estimated and the real camera poses are $\hat{P} \in SE(3)$ and $P \in SE(3)$, respectively. Typically, the translation and rotation parts of the RPE are separated.

$$\text{trans}_{error}^i = \left(\frac{1}{M} \sum_1^M \left\| \text{trans}(E_i) \right\|^2 \right)^{\frac{1}{2}} \quad (11)$$

As for the rotation component, we use the mean error approach:

$$\text{rot}_{error}^i = \frac{1}{M} \sum_1^M \angle(\text{rot}(E_i^\Delta)) \quad (12)$$

Averaging both the translation and rotation components of SLAM systems is a sensible approach for evaluating these systems.

TABLE I. COMPARING THE RMSE OF THE SEQUENCES 01 AND 03 USING DIFFERENT METHODS

Method \Sequences	Sequence 01				Sequence 03			
	$tran_{error}$	rot_{error}	$RPE(m)$	$RPE(^{\circ})$	$tran_{error}$	rot_{error}	$RPE(m)$	$RPE(^{\circ})$
ORB-SLAM2(without LC)	107.57	0.89	2.970	0.098	0.97	0.19	0.031	0.055
DF-VO(Mono-SC Train)	66.98	17.04	1.281	0.725	2.67	0.50	0.030	0.038
VISO2	61.36	7.68	1.413	0.432	30.21	2.21	0.226	0.157
SfM-Learner	22.41	2.79	0.660	0.133	12.56	4.52	0.077	0.158
Depth-VO-Feat	23.78	1.75	0.547	0.133	15.76	10.62	0.168	0.308
Our Method	21.53	1.64	0.471	0.125	6.81	2.23	0.201	0.1689

TABLE II. COMPARING THE RMSE OF THE SEQUENCES 09 AND 10 USING DIFFERENT METHODS

Method \Sequences	Sequence 09				Sequence 10			
	$tran_{error}$	rot_{error}	$RPE(m)$	$RPE(^{\circ})$	$tran_{error}$	rot_{error}	$RPE(m)$	$RPE(^{\circ})$
ORB-SLAM2(Without LC)	9.30	0.26	0.128	0.061	2.57	0.32	0.045	0.065
DF-VO(Mono)	2.47	0.30	0.055	0.037	1.96	0.31	0.047	0.042
VISO2	18.06	1.25	0.284	0.125	26.10	3.26	0.442	0.154
SfM-Learner	11.32	4.07	0.103	0.159	15.25	4.06	0.118	0.171
Depth-VO-Feat	11.89	3.60	0.164	0.233	12.82	3.41	0.159	0.246
Our Method	3.41	1.42	0.094	0.147	16.25	7.82	0.148	0.187

$$\angle S := \arccos\left(\frac{\text{tr}(S) - 1}{2}\right) \quad (13)$$

We performed a qualitative experiment comparing this approach to visual odometry with numerous cutting-edge VO techniques to evaluate its applicability for estimating scale, including traditional monocular, stereo approaches, and learning approaches for monocular odometry such as ORB-SLAM2[64], DF-VO[63], VISO2[65], SfM-Learner[54], and Depth-VO-Feat[66]. Conventional monocular VO techniques necessitate posture with a prior knowledge of ground truth and are unable to recover the absolute scale. The global loop-closure detection of the ORB-SLAM2 has been deactivated in order to create an equivalent comparison. The keyframe trajectories of the ORB-SLAM2 are matched to ground truth via similarity transformation because it cannot retrieve the absolute scale.

On the KITTI odometry dataset's sequences 01, 03, 09, and 10, respectively, Fig. 6 compares the trajectories obtained by our approach to the ground truth trajectories. How closely the trajectory produced by our technique and the ground truth in the numbers 01, 03, 09, and 10 correspond. The metrics were computed using the KITTI evaluation toolkit for the KITTI sequences with ground truth. The quantitative results are shown in Tables I and II, and the best metric values are denoted by bolded values. Our strategy produced good results and was comparable with the other approaches, but the DF-VO remains the accurate framework not only for the four trajectories but for all trajectories of KITTI-Dataset. Additionally, in the various four sequences, the $terr$ and $rerr$ both produced good results that were greater to those of some other methods in the

four trajectory. For rotation and translation RPEs, our model performed significantly enough.

This shows that to reduce scale drift problems, it is essential to have a depth map computed by a high accuracy and precise model in scenarios where depth-map estimation is used to compute the scale. We showed that a transformer-based method can produce results that are comparable to or even better than CNN-based techniques when used as a monocular visual odometry system component.

VI. CONCLUSION

In this paper, we present a method for estimating scale in visual odometry using a dense prediction transformer model. Due to our model's high performance in estimating depth maps from a monocular camera, scale drifts were reduced in multiple visual odometry sequences of the KITTI dataset. As a result of our experimental results on the KITTI odometry benchmark, we are confident that our proposed method is not only accurate enough but also shows a similar result to state-of-the-art approaches.

While data fusion-based visual approaches provide the highest accuracy of localization, they have some limitations, such as being computationally expensive. Real-time operation on resource-constrained systems is still possible if implemented efficiently. As a part of our future work, we will concentrate on developing a real-time integration of GNSS, IMU, and Lidar data using one of the extensions of the Kalman filter. In order to optimize the time consumption of low-cost embedded system implementation without losing accuracy.

ACKNOWLEDGMENT

We would like to express our gratitude to the Moroccan National Center for Scientific and Technical Research (CNRST) for its encouragement (grant number: 37 UM5R2022) during the period June 2022 to April 2023.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

- [1] PHAM, Minh et XIONG, Kaiqi. A survey on security attacks and defense techniques for connected and autonomous vehicles. *Computers & Security*, 2021, vol. 109, p. 102269.
- [2] MANUEL, Melvin P., FAIED, Mariam, KRISHNAN, Mohan, et al. Robot Platooning Strategy for Search and Rescue Operations. *Intelligent Service Robotics*, 2022, vol. 15, no 1, p. 57-68.
- [3] KUMAR, Amit, OJHA, Aparajita, YADAV, Sonal, et al. Real-time interception performance evaluation of certain proportional navigation based guidance laws in aerial ground engagement. *Intelligent Service Robotics*, 2022, vol. 15, no 1, p. 95-114.
- [4] TIAN, Ying, YAO, Qiangqiang, WANG, Chengqiang, et al. Switched model predictive controller for path tracking of autonomous vehicle considering rollover stability. *Vehicle System Dynamics*, 2022, vol. 60, no 12, p. 4166-4185.
- [5] LI, Qingqing, QUERALTA, Jorge Peña, GIA, Tuan Nguyen, et al. Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. *Unmanned Systems*, 2020, vol. 8, no 03, p. 229-237.
- [6] SABIHA, Ahmed D., KAMEL, Mohamed A., SAID, Ehab, et al. Real-time path planning for autonomous vehicle based on teaching-learning-based optimization. *Intelligent Service Robotics*, 2022, vol. 15, no 3, p. 381-398.
- [7] MENG, Lingbo, YE, Chao, et LIN, Weiyang. A tightly coupled monocular visual lidar odometry with loop closure. *Intelligent Service Robotics*, 2022, vol. 15, no 1, p. 129-141.
- [8] YEONG, De Jong, VELASCO-HERNANDEZ, Gustavo, BARRY, John, et al. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 2021, vol. 21, no 6, p. 2140.
- [9] DU, Hao, WANG, Wei, XU, Chaowen, et al. Real-time onboard 3D state estimation of an unmanned aerial vehicle in multi-environments using multi-sensor data fusion. *Sensors*, 2020, vol. 20, no 3, p. 919.
- [10] XU, Xiaobin, ZHANG, Lei, YANG, Jian, et al. A review of multi-sensor fusion slam systems based on 3D LIDAR. *Remote Sensing*, 2022, vol. 14, no 12, p. 2835.
- [11] MARKOVIĆ, Lovro, KOVAČ, Marin, MILIJAS, Robert, et al. Error state extended Kalman filter multi-sensor fusion for unmanned aerial vehicle localization in GPS and magnetometer denied indoor environments. In : 2022 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2022. p. 184-190.
- [12] ZHOU, Yi, GALLEGO, Guillermo, et SHEN, Shaojie. Event-based stereo visual odometry. *IEEE Transactions on Robotics*, 2021, vol. 37, no 5, p. 1433-1450.
- [13] ABOUZAHIR, Mohamed, ELOUARDI, Abdelhafid, LATIF, Rachid, et al. Embedding SLAM algorithms: Has it come of age?. *Robotics and Autonomous Systems*, 2018, vol. 100, p. 14-26.
- [14] FERRERA, Maxime, EUDES, Alexandre, MORAS, Julien, et al. OV² SLAM: A Fully Online and Versatile Visual SLAM for Real-Time Applications. *IEEE Robotics and Automation Letters*, 2021, vol. 6, no 2, p. 1399-1406.
- [15] CHENG, Jun, ZHANG, Liyan, CHEN, Qihong, et al. A review of visual SLAM methods for autonomous driving vehicles. *Engineering Applications of Artificial Intelligence*, 2022, vol. 114, p. 104992.
- [16] CAMPOS, Carlos, ELVIRA, Richard, RODRÍGUEZ, Juan J. Gómez, et al. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics*, 2021, vol. 37, no 6, p. 1874-1890.
- [17] YU, Zhelin, ZHU, Lidong, et LU, Guoyu. Tightly-coupled Fusion of VINS and Motion Constraint for Autonomous Vehicle. *IEEE Transactions on Vehicular Technology*, 2022.
- [18] QIN, Tong, LI, Peiliang, et SHEN, Shaojie. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 2018, vol. 34, no 4, p. 1004-1020.
- [19] SATTLER, Torsten, TORII, Akihiko, SIVIC, Josef, et al. Are large-scale 3d models really necessary for accurate visual localization?. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 1637-1646.
- [20] YIN, Xiaochuan, WANG, Xiangwei, DU, Xiaoguo, et al. Scale recovery for monocular visual odometry using depth estimated with deep convolutional neural fields. In : Proceedings of the IEEE international conference on computer vision. 2017. p. 5870-5878.
- [21] ZHANG, Hui, WANG, Xiangwei, YIN, Xiaochuan, et al. Geometry-Constrained Scale Estimation for Monocular Visual Odometry. *IEEE Transactions on Multimedia*, 2021.
- [22] ÖLMEZ, Burhan et TUNCER, Temel Engin. Metric scale and angle estimation in monocular visual odometry with multiple distance sensors. *Digital Signal Processing*, 2021, vol. 117, p. 103148.
- [23] TIAN, Rui, ZHANG, Yunzhou, ZHU, Delong, et al. Accurate and robust scale recovery for monocular visual odometry based on plane geometry. In : 2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021. p. 5296-5302.
- [24] LI, Ruihao, WANG, Sen, et GU, Dongbing. Ongoing evolution of visual slam from geometry to deep learning: Challenges and opportunities. *Cognitive Computation*, 2018, vol. 10, no 6, p. 875-889.
- [25] ARSHAD, Saba et KIM, Gon-Woo. Role of deep learning in loop closure detection for visual and lidar slam: A survey. *Sensors*, 2021, vol. 21, no 4, p. 1243.
- [26] CHAPLOT, Devendra Singh, GANDHI, Dhiraj, GUPTA, Saurabh, et al. Learning to explore using active neural slam. *arXiv preprint arXiv:2004.05155*, 2020.
- [27] DUAN, Chao, JUNGINGER, Steffen, HUANG, Jiahao, et al. Deep learning for visual SLAM in transportation robotics: a review. *Transportation Safety and Environment*, 2019, vol. 1, no 3, p. 177-184.
- [28] PEDRAZA, Luis, RODRIGUEZ-LOSADA, Diego, MATIA, Fernando, et al. Extending the limits of feature-based SLAM with B-splines. *IEEE Transactions on Robotics*, 2009, vol. 25, no 2, p. 353-366.
- [29] ENGEL, Jakob, STÜCKLER, Jörg, et CREMERS, Daniel. Large-scale direct SLAM with stereo cameras. In : 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2015. p. 1935-1942.
- [30] SILVEIRA, Geraldo, MALIS, Ezio, et RIVES, Patrick. An efficient direct approach to visual SLAM. *IEEE transactions on robotics*, 2008, vol. 24, no 5, p. 969-979.
- [31] STRASDAT, Hauke, MONTIEL, J., et DAVISON, Andrew J. Scale drift-aware large scale monocular SLAM. *Robotics: Science and Systems VI*, 2010, vol. 2, no 3, p. 7.
- [32] STRASDAT, Hauke, MONTIEL, J., et DAVISON, Andrew J. Scale drift-aware large scale monocular SLAM. *Robotics: Science and Systems VI*, 2010, vol. 2, no 3, p. 7.
- [33] KRIZHEVSKY, Alex, SUTSKEVER, Ilya, et HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, vol. 60, no 6, p. 84-90.
- [34] KONRAD, Janusz, WANG, Meng, et ISHWAR, Prakash. 2d-to-3d image conversion by learning depth from examples. In : 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2012. p. 16-22.
- [35] KARSCH, Kevin, LIU, Ce, et KANG, Sing Bing. Depth extraction from video using non-parametric sampling. In : European conference on computer vision. Springer, Berlin, Heidelberg, 2012. p. 775-788.
- [36] MUR-ARTAL, Raul, MONTIEL, Jose Maria Martinez, et TARDOS, Juan D. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 2015, vol. 31, no 5, p. 1147-1163.
- [37] GAO, Yuan et YUILLE, Alan L. Symmetric non-rigid structure from motion for category-specific object structure estimation. In : European Conference on Computer Vision. Springer, Cham, 2016. p. 408-424.

- [38] GAO, Yuan et YUILLE, Alan L. Exploiting symmetry and/or manhattan properties for 3d object structure estimation from single and multiple images. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 7408-7417.
- [39] MA, Jiayi, ZHOU, Huabing, ZHAO, Ji, et al. Robust feature matching for remote sensing image registration via locally linear transforming. IEEE Transactions on Geoscience and Remote Sensing, 2015, vol. 53, no 12, p. 6469-6481.
- [40] EIGEN, David, PUHRSCHE, Christian, et FERGUS, Rob. Depth map prediction from a single image using a multi-scale deep network. Advances in neural information processing systems, 2014, vol. 27.
- [41] EIGEN, David et FERGUS, Rob. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In : Proceedings of the IEEE international conference on computer vision. 2015. p. 2650-2658.
- [42] LAINA, Iro, RUPPRECHT, Christian, BELAGIANNIS, Vasileios, et al. Deeper depth prediction with fully convolutional residual networks. In : 2016 Fourth international conference on 3D vision (3DV). IEEE, 2016. p. 239-248.
- [43] LI, Jun, KLEIN, Reinhard, et YAO, Angela. A two-streamed network for estimating fine-scaled depth maps from single rgb images. In : Proceedings of the IEEE International Conference on Computer Vision. 2017. p. 3372-3380.
- [44] YAN, Han, ZHANG, Shunli, ZHANG, Yu, et al. Monocular depth estimation with guidance of surface normal map. Neurocomputing, 2018, vol. 280, p. 86-100.
- [45] LI, Bo, SHEN, Chunhua, DAI, Yuchao, et al. Depth and surface normal estimation from monocular images using regression on deep features and hierarchical crfs. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1119-1127.
- [46] ROY, Anirban et TODOROVIC, Sinisa. Monocular depth estimation using neural regression forest. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 5506-5514.
- [47] LIU, Fayao, SHEN, Chunhua, et LIN, Guosheng. Deep convolutional neural fields for depth estimation from a single image. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 5162-5170.
- [48] GARG, Ravi, BG, Vijay Kumar, CARNEIRO, Gustavo, et al. Unsupervised cnn for single view depth estimation: Geometry to the rescue. In : European conference on computer vision. Springer, Cham, 2016. p. 740-756.
- [49] ILG, Eddy, MAYER, Nikolaus, SAIKIA, Tonmoy, et al. Flownet 2.0: Evolution of optical flow estimation with deep networks. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2462-2470.
- [50] YU, Jason J., HARLEY, Adam W., et DERPANIS, Konstantinos G. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In : European Conference on Computer Vision. Springer, Cham, 2016. p. 3-10.
- [51] GODARD, Clément, MAC AODHA, Oisín, et BROSTOW, Gabriel J. Unsupervised monocular depth estimation with left-right consistency. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 270-279.
- [52] KUZNIETSOV, Yevhen, STUCKLER, Jorg, et LEIBE, Bastian. Semi-supervised deep learning for monocular depth map prediction. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 6647-6655.
- [53] HUA, Yan et TIAN, Hu. Depth estimation with convolutional conditional random field network. Neurocomputing, 2016, vol. 214, p. 546-554.
- [54] ZHOU, Tinghui, BROWN, Matthew, SNAVELY, Noah, et al. Unsupervised learning of depth and ego-motion from video. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 1851-1858.
- [55] YIN, Zhichao et SHI, Jianping. Geonet: Unsupervised learning of dense depth, optical flow and camera pose. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 1983-1992.
- [56] LUO, Yue, REN, Jimmy, LIN, Mude, et al. Single view stereo matching. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. p. 155-163.
- [57] VASWANI, Ashish, SHAZEER, Noam, PARMAR, Niki, et al. Attention is all you need. Advances in neural information processing systems, 2017, vol. 30.
- [58] DEVLIN, Jacob, CHANG, Ming-Wei, LEE, Kenton, et al. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [59] RADFORD, Alec, NARASIMHAN, Karthik, SALIMANS, Tim, et al. Improving language understanding by generative pre-training. 2018.
- [60] DOSOVITSKIY, Alexey, BEYER, Lucas, KOLESNIKOV, Alexander, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020.
- [61] RANFTL, René, BOCHKOVSKIY, Alexey, et KOLTUN, Vladlen. Vision transformers for dense prediction. In : Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. p. 12179-12188.
- [62] ROSTEN, Edward, PORTER, Reid, et DRUMMOND, Tom. Faster and better: A machine learning approach to corner detection. IEEE transactions on pattern analysis and machine intelligence, 2008, vol. 32, no 1, p. 105-119.
- [63] ZHAN, Huangying, WEERASEKERA, Chamara Saroj, BIAN, Jia-Wang, et al. DF-VO: What Should Be Learnt for Visual Odometry?. arXiv preprint arXiv:2103.00933, 2021.
- [64] MUR-ARTAL, Raul et TARDÓS, Juan D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. IEEE transactions on robotics, 2017, vol. 33, no 5, p. 1255-1262.
- [65] GEIGER, Andreas, ZIEGLER, Julius, et STILLER, Christoph. Stereoscan: Dense 3d reconstruction in real-time. In : 2011 IEEE intelligent vehicles symposium (IV). Ieee, 2011. p. 963-968.
- [66] ZHAN, Huangying, GARG, Ravi, WEERASEKERA, Chamara Saroj, et al. Unsupervised learning of monocular depth estimation and visual odometry with deep feature reconstruction. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 340-349.

Enhancing Precision in Lung Cancer Diagnosis Through Machine Learning Algorithms

Nasareenbanu Devihosur¹, Ravi Kumar M G²

Research Scholar, School of Electronics and Communication Engineering, REVA University, Bangalore-560064¹

Department of AI/ML, KNS Institute of Technology, Bengaluru-570064¹

Nagarjuna College of Engineering and Technology, Devanahalli, Bengaluru, Karnataka-562110²

Abstract—Lung cancer continues to pose a significant threat worldwide, leading to high cancer-related mortality rates and underscoring the urgent need for improved early diagnosis approaches. Despite the valuable technology currently employed for lung cancer diagnosis, some limitations hinder timely and accurate diagnoses, resulting in delayed treatment and unfavorable outcomes. In this research, we propose a comprehensive methodology that harnesses the power of various machine learning algorithms, including Logistic Regression, Gradient Boost, LGBM, and Support Vector Machine, to address these challenges and improve patient care. These algorithms have been thoughtfully chosen for their ability to effectively handle the complexity of lung cancer data and enable accurate classification and prediction of cases. By leveraging these advanced techniques, our methodology aims to enhance the efficiency and accuracy of lung cancer diagnosis, enabling earlier interventions and tailored treatment plans that can significantly impact patient outcomes and quality of life. Through rigorous assessments conducted on benchmark datasets and real-world cases, our study has yielded promising results. Random Forest achieved an impressive accuracy of 97%, showcasing its ability to effectively capture complex patterns and features within the lung cancer dataset. By pushing the boundaries of medical innovation and precision medicine, we envision a future where machine learning algorithms seamlessly integrate into healthcare systems, leading to personalized and efficient care for lung cancer patients.

Keywords—Lung cancer diagnosis; machine learning; precision medicine

I. INTRODUCTION

Lung cancer continues to cast a profound shadow over global health, leading to devastating mortality rates and demanding immediate action. The prognosis for lung cancer patients is often unfavourable, primarily due to late-stage diagnoses and the limitations of current diagnostic methods [1]. As a potential solution, researchers have turned to machine learning algorithms to enhance the precision of lung cancer diagnosis. Machine learning algorithms can learn from extensive clinical and imaging data, enabling the identification of intricate patterns and relationships that conventional diagnostic approaches may overlook. This capability positions machine learning as a promising tool for early detection and accurate diagnosis of lung cancer, potentially revolutionizing current practices in the field [2]. Fig. 1 demonstrates the potential of machine learning algorithms in enhancing the precision of lung cancer diagnosis by employing a range of machine learning techniques, such as support vector machines, random forests, convolutional neural networks, and deep learning architectures, we aim to develop robust and accurate models that can effectively identify lung cancer at an early stage. The utilization of

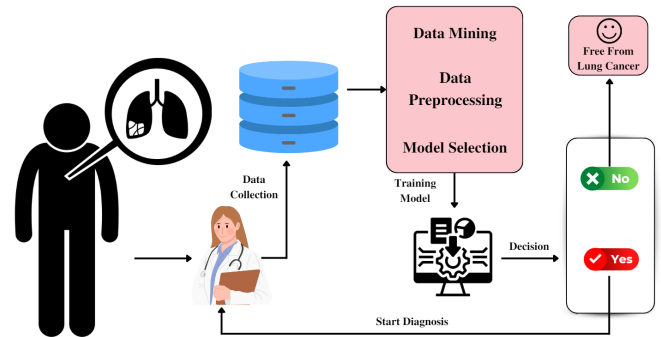


Fig. 1. Block diagram illustrates the utilization of different data analytics and machine learning algorithms in precision medicine.

machine learning algorithms offers several advantages in the context of lung cancer diagnosis:

- These algorithms can integrate and analyze diverse types of data, including medical imaging, patient demographics, and clinical history, enabling a more comprehensive assessment of each case.
- Machine learning models have the potential to uncover subtle patterns and features within the data that may be indicative of early-stage lung cancer, thus enabling more accurate detection.
- Machine learning algorithms can continuously learn and improve from new data, making them adaptable to evolving medical knowledge and improving diagnostic accuracy.

Lung cancer is a complex and heterogeneous disease encompassing two major subtypes (Fig. 2Prevalence of NSCLC and SCLC of lung cancer.): Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC). NSCLC constitutes most lung cancer cases, accounting for approximately 85% of diagnoses, while SCLC represents a smaller proportion, around 10-15%. Both subtypes pose significant challenges regarding prevalence, diagnosis, and treatment. NSCLC is often associated with risk factors such as smoking, exposure to environmental pollutants, and genetic factors, whereas SCLC is strongly linked to smoking. Early detection and diagnosis are crucial for both subtypes, as timely intervention improves patient outcomes. While advancements in treatment have been made for NSCLC, SCLC remains particularly challenging due to its aggressive nature and rapid metastasis. Targeted therapies

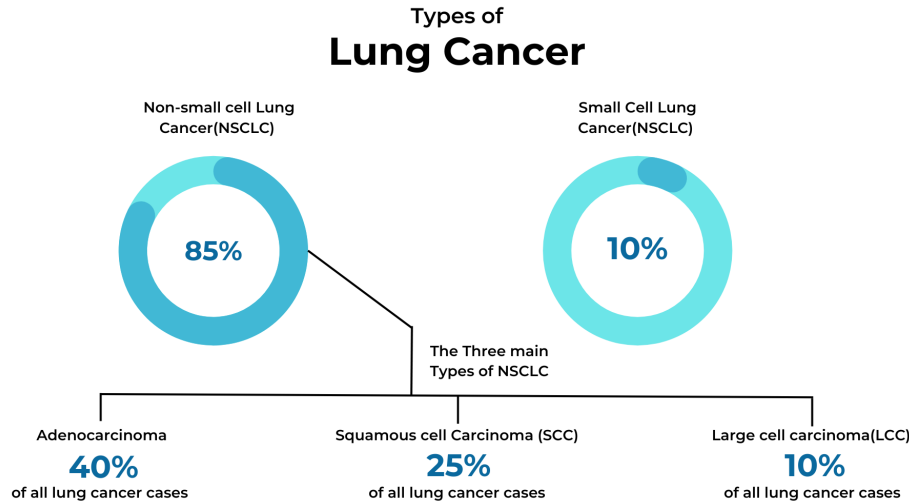


Fig. 2. Prevalence of NSCLC and SCLC of lung cancer.

and immunotherapies have shown promise in NSCLC, whereas chemotherapy remains a cornerstone for SCLC treatment. Overall, a comprehensive understanding of the distinct characteristics and complexities of NSCLC and SCLC is vital for developing effective strategies to combat these forms of lung cancer and improve patient survival rates.

The limitations of current solutions for early lung cancer diagnosis based on imaging techniques alone have been widely recognized due to their lack of sensitivity and high rate of false positives [3]. Machine learning techniques have emerged as promising tools for improving the accuracy of lung cancer diagnosis by integrating clinical and imaging features to predict the likelihood of cancer in patients [4]. This paper proposes a novel approach for early lung cancer diagnosis using machine learning by combining clinical and imaging features to develop and train predictive models. Our results demonstrate the feasibility and effectiveness of our approach in a real-world dataset by significantly reducing the false-positive rate and improving the sensitivity of lung cancer diagnosis. Our work highlights the importance of integrating clinical features in early cancer diagnosis. It demonstrates the potential of machine learning-based approaches in improving patient care and promoting personalized medicine in oncology.

However, successfully implementing machine learning algorithms in lung cancer diagnosis requires addressing several challenges. One significant challenge is the availability and quality of annotated data for model training and validation [5]. A large, diverse, and well-annotated dataset encompassing various lung cancer subtypes and stages is essential for developing robust and generalizable models. Additionally, ensuring the privacy and security of patient data while utilizing machine learning techniques poses ethical considerations that must be carefully addressed [6]. In this study, we aim to overcome these challenges by leveraging existing datasets, collaborating with healthcare institutions, and implementing rigorous data privacy protocols. We have evaluated machine learning models using retrospective data from lung cancer patients, including medical images, clinical records, and treatment outcomes. The

performance of these models will be rigorously assessed using appropriate evaluation metrics, such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC). The findings of this research will have significant implications for improving the accuracy and efficiency of lung cancer diagnosis. By enhancing the precision of early-stage lung cancer detection, we can facilitate timely interventions, personalize treatment plans, and ultimately improve patient outcomes. Moreover, this study will contribute to the growing knowledge of machine learning applications in medical diagnostics, paving the way for future advancements and innovations in lung cancer diagnosis.

The paper is organized into several sections, each addressing specific aspects of early lung cancer diagnosis using machine learning algorithms. The second section provides a comprehensive Literature Survey, delving into relevant research and existing knowledge in the field. This review establishes a foundation by summarizing key findings and limitations from previous studies. Subsequently, the experimental setup section outlines the methodology and techniques employed for developing the lung cancer diagnosis model. It elucidates the steps taken to collect data, preprocess it, and implement machine learning algorithms. Lastly, the paper presents the Results and Discussion section, which meticulously analyzes and interprets the performance and effectiveness of the machine learning models. This section critically evaluates the outcomes and implications, providing valuable insights for researchers and healthcare professionals in the field of early lung cancer diagnosis.

II. BACKGROUND

Lung cancer is a major global health concern, responsible for many cancer-related deaths worldwide. According to the World Health Organization (WHO), lung cancer accounted for approximately 2.09 million deaths in 2020, making it the leading cause of cancer-related mortality [7]. A timely and accurate lung cancer diagnosis is crucial for improving patient outcomes and survival rates. However, the current diagnostic

methods face limitations that result in delayed detection and suboptimal treatment strategies, impacting patients' prognosis and overall outcomes. Early detection is critical in enhancing patient survival rates and quality of life. Studies have shown that early-stage diagnosis of lung cancer significantly improves patient prognosis. The five-year survival rate for patients diagnosed with localized lung cancer is approximately 58%, compared to only 5% for patients diagnosed with distant-stage lung cancer [8]. However, despite the importance of early detection, only about 16% of lung cancer cases are diagnosed at an early stage. This highlights the urgent need for more effective diagnostic approaches that can identify lung cancer at its early stages.

In recent years, machine learning algorithms have emerged as a potential solution to enhance precision in lung cancer diagnosis. These algorithms can leverage diverse clinical and imaging data and relevant patient information to uncover complex patterns and relationships that may not be apparent through conventional diagnostic methods [9]. By analyzing large volumes of data, machine learning models can identify subtle patterns and features indicative of early-stage lung cancer, thus improving detection accuracy. Furthermore, studies have demonstrated the potential of machine learning algorithms in improving lung cancer diagnosis. Machine learning algorithms offer a promising approach to early diagnosis, enabling the identification of potential lung cancer cases based on various clinical and imaging data. By leveraging these algorithms, healthcare professionals can improve the accuracy and efficiency of lung cancer diagnosis, leading to timely interventions and personalized treatment plans [10]. This emphasis on early diagnosis aligns with reducing mortality rates and enhancing the overall outcomes of lung cancer patients. Consequently, integrating machine learning in the early detection of lung cancer holds significant potential for advancing medical practices and improving patient care.

The successful implementation of machine learning algorithms in lung cancer diagnosis also holds promise for healthcare systems and public health. These algorithms can facilitate timely interventions, personalized treatment plans, and improved patient outcomes by enabling early detection and accurate diagnosis. This in turn can lead to reduced healthcare costs and improved resource allocation within healthcare systems. Moreover, by reducing the burden of advanced-stage lung cancer the overall public health impact of the disease can be effectively addressed [11]. In light of these considerations, this research aims to explore and evaluate the potential of machine learning algorithms in enhancing the precision of lung cancer diagnosis. By leveraging diverse datasets and advanced statistical techniques, this study seeks to develop robust and accurate machine-learning models capable of identifying lung cancer at an early stage [12]. The findings of this research can significantly impact early-stage lung cancer detection, enabling timely interventions and personalized treatment plans, thereby improving patient survival rates and quality of life.

Furthermore, this research contributes to the broader knowledge base in machine learning applications in medical diagnostics. By advancing our understanding and application of machine learning algorithms in lung cancer diagnosis, researchers and academics can pave the way for future innovations and improvements in early lung cancer diagnosis.

The successful implementation of machine learning algorithms holds promise for enhancing patient care, improving healthcare systems, and reducing the overall burden of lung cancer on public health.

III. LITERATURE SURVEY

Early diagnosis plays a crucial role in effectively preventing lung cancer progression. Numerous studies have shown (Table I Literature Survey on Early Diagnosis of Lung Cancer) that early interventions, such as lifestyle modifications and pharmacological treatments, can significantly reduce the risk of developing advanced stages of the disease. Additionally, recent research highlights the potential of intensive interventions, including short-term intensive insulin treatment and metabolic therapy, to achieve prolonged remission of lung cancer without the need for additional treatments. Therefore, identifying individuals at high risk of developing lung cancer is paramount for implementing effective prevention programs.

In a study by Ardila et al. (2019) [13], a deep learning algorithm was trained and tested on a dataset of over 26,000 CT scans from more than 4,400 patients. The algorithm identified lung nodules with an accuracy of 94.4%, outperforming radiologists in the same task. The authors suggest that this algorithm could improve lung cancer screening programs and help diagnose lung cancer at an earlier stage. Another study by Lia0 et al. (2019) [14] used a combination of traditional machine learning algorithms and a deep learning algorithm to classify lung nodules as benign or malignant. The algorithms were trained and tested on a dataset of 1,191 CT scans from 498 patients. The deep learning algorithm had an accuracy of 91.1%, outperforming the traditional machine learning algorithms. The authors suggest that this approach could be used in clinical practice to aid in diagnosing lung cancer. In a review article by Wang et al. (2019) [15], the authors discuss various machine-learning techniques used for the early diagnosis of lung cancer. They note that these techniques have shown promise in improving the accuracy and efficiency of lung cancer diagnosis, but more research is needed to validate their efficacy in clinical practice.

Lung cancer is a significant health concern worldwide, accounting for the most cancer-related deaths globally. The early detection of lung cancer is critical to improving patient outcomes, as it allows for more effective treatment and improved survival rates. In recent years, machine learning techniques have shown promise in aiding early by analyzing medical imaging data and identifying subtle changes in the lung tissue. Naik et al. (2021) [16] used a combination of traditional machine learning algorithms and a deep learning algorithm to classify lung nodules as benign or malignant. Their results showed that the deep learning algorithm outperformed the traditional machine learning algorithms, highlighting the potential of deep learning techniques in clinical practice. Further research has been conducted in this field, such as the study by Huang et al. (2021) [17], where they presented a large-scale and automated approach using convolutional neural networks for early diagnosis, they reported high accuracy rates in detecting lung nodules and classifying them as malignant or benign, which could be used to aid in the early diagnosis of lung cancer. Saleh et al. (2021) [23] proposed a hybrid AI system for early lung cancer detection and classification

TABLE I. LITERATURE SURVEY ON EARLY DIAGNOSIS OF LUNG CANCER

Study	Year	Methodology	Key Findings	Limitations
Wang et al. [18]	2017	Machine Learning (Random Forest)	Achieved 95% accuracy in early lung cancer detection using radiomics features	Small sample size, limited external validation
Hosny et al.[19]	2018	Deep Learning (Convolutional Neural Networks)	Developed a model with 90% sensitivity and 92% specificity in detecting lung cancer from CT scans	Reliance on annotated data, potential overfitting
Singal et al. [20]	2019	Biomarker Analysis	Identified a panel of circulating microRNAs with high sensitivity and specificity for early lung cancer diagnosis	Limited sample diversity, need for further validation
Mehta et al. [21]	2020	Hybrid Model (Machine Learning + Imaging)	Combined radiomic features and clinical variables to achieve 87% accuracy in distinguishing malignant lung nodules from benign ones	Limited interpretability, potential bias in feature selection
Gürsoy et al. [22]	2021	Artificial Intelligence (AI) Based System	Developed an AI system with 96% accuracy in classifying lung nodules as malignant or benign based on CT images	Limited generalization to diverse datasets, need for real-world evaluation

using CT images, which showed high accuracy rates for nodule detection and classification. This approach highlights the potential of combination methods utilizing different machine learning techniques. Lu et al. (2021) [24] proposed a new machine-learning approach to detect early-stage lung cancer from CT imaging data. Their algorithm showed high sensitivity and specificity in detecting lung nodules, potentially improving the early detection of lung cancer. Gu et al. [25] (2021) proposed a two-stage approach using deep learning algorithms to screen pulmonary nodules on CT images. Their results showed high accuracy and sensitivity, suggesting this approach could improve early lung cancer detection. Wu et al. (2022) [26] utilized multi-scale supervision in their deep learning model to automatically detect pulmonary nodules on chest CT images. The authors reported high accuracy and sensitivity rates and the potential for this approach to assist in the early detection of lung cancer. Huang et al. (2023) [27] proposed a hybrid approach using deep learning algorithms and radiomics analysis for the automated diagnosis and classification of lung cancer. Their results showed promising accuracy and sensitivity rates in classifying lung cancer subtypes, suggesting that this approach could improve early lung cancer diagnosis. Huh et al. (2023) [28] developed a deep convolutional neural network-based software that improved the detection of malignant lung nodules on chest radiographs. Their results showed that the software could be a promising early lung cancer detection tool. Lv et al. (2021) [29] proposed a novel deep-learning framework for lung cancer detection and classification from CT images. Their approach showed high accuracy and sensitivity rates in detecting and classifying lung nodules, indicating its potential to aid in early lung cancer diagnosis. Bilal et al. (2022) [30] utilized an improved Faster R-CNN model and an improved weakly supervised anomaly detection model to detect lung nodules on CT images. Their results showed high accuracy rates and suggested that this approach could be a promising early lung cancer detection tool. Liu et al. (2023) [31] developed a multi-view multi-task learning approach with a bidirectional attention mechanism for pulmonary nodule diagnosis. Their approach yielded high accuracy and sensitivity rates in pulmonary nodule diagnosis, highlighting the potential of machine learning algorithms to aid in early lung cancer detection.

IV. EXPERIMENTAL SETUP

To evaluate the effectiveness of our proposed methodology (Fig. 4Block representation of the proposed model.) for enhancing precision in lung cancer diagnosis through machine

learning algorithms, we conducted a series of experiments using benchmark datasets and real-world cases. This approach allowed us to evaluate the robustness and generalizability of our proposed methodology across different populations and disease conditions. The following outlines the key components of our experimental setup:

A. About Dataset

The dataset used in this study is collected from National Cancer Institute and consists of lung cancer data, providing a valuable resource for our research on applying AI/ML algorithms to improve the diagnosis of lung cancer. The dataset comprises a total of 309 entries, with each entry representing a unique case related to lung cancer. Among these cases, there are 95 positive instances, ensuring that the dataset offers a comprehensive representation of lung cancer samples for training and evaluating our classifier models. Each instance

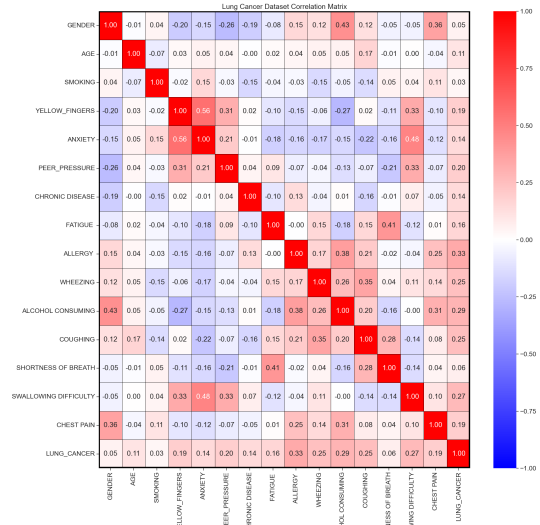


Fig. 3. Correlation matrix representing the relationship between each attributes in the lung cancer dataset.

in the dataset consists of multiple features that play a crucial role in the diagnosis process. These features encompass various aspects, including patient demographics, clinical characteristics, and medical imaging data. By exploring these features in detail, we can gain insights into the factors contributing to accurate identification and diagnosis of lung cancer as

demonstrated in the correlation matrix (Fig. 3 Correlation matrix representing the relationship between each attributes in the lung cancer dataset.). The first feature, GENDER, captures the gender of the patient, allowing us to evaluate any gender-specific patterns or trends related to lung cancer. Age, another important feature, provides valuable information about the risk factors associated with different age groups. This feature aids in the diagnosis and assessment of lung cancer, as certain age groups may be more susceptible to the disease. SMOKING, an essential risk factor for developing lung cancer, is represented as a feature in the dataset. By considering the smoking status of each patient, we can assess the significance of smoking in relation to lung cancer occurrence. Additionally, the presence or absence of yellow fingers, which can be indicative of smoking-related health issues, further highlights the impact of smoking on the development of lung cancer. Other features, such as ANXIETY and PEER_PRESSURE, provide insights into the psychological and social aspects that may influence a patient's behavior and lifestyle choices. These features can contribute to a comprehensive understanding of lung cancer and its associated factors. Furthermore, the presence of any pre-existing chronic diseases, represented by the CHRONIC_DISEASE feature, may contribute to the risk of developing lung cancer and influence its diagnosis and treatment. Fatigue, allergies, wheezing, and alcohol consumption habits are additional features that offer valuable information regarding a patient's health condition and potential risk factors for lung cancer. Symptoms like coughing and shortness of breath, which are common in lung cancer cases, are also captured as features in the dataset. Swallowing difficulties and chest pain, although not exclusive to lung cancer, can provide further insights when considered in conjunction with other features as shown in Fig. 5 Exploratory data analysis for lung cancer diagnosis.. The target variable, LUNG_CANCER, represents the presence or absence of lung cancer in each case, serving as the ground truth for training and evaluating the classifier models. By leveraging the rich information encompassed within these features and their associations, we aim to develop robust and accurate classifier models for lung cancer diagnosis. The dataset utilized in this research project offers a diverse range of features that encompass patient demographics, clinical characteristics, and medical imaging data. Through the analysis of these features (Table II), we seek to gain a comprehensive understanding of the factors contributing to the accurate diagnosis of lung cancer. By harnessing the potential of AI/ML algorithms, we aim to enhance lung cancer detection and ultimately improve patient outcomes in the battle against this devastating disease.

B. Split Dataset

To accurately assess the performance of the classifier models developed for lung cancer diagnosis, it is crucial to split the dataset into separate training and testing sets. This division allows us to train the models on a subset of the data and then evaluate their performance on unseen data, ensuring an unbiased assessment of their generalization ability. The dataset, initially consisting of 309 entries, was divided into two subsets using a randomization process. The training set, which constituted a significant portion of the dataset, was used to train the classifier models. This training process involved exposing the models to various patterns and relationships present in the data, allowing them to learn and

make predictions based on the provided features. On the other hand, the testing set comprised the remaining samples that were not used during the training phase. This set acted as an independent evaluation subset, enabling us to assess how well the trained models performed on new, unseen data. By evaluating the models' performance on the testing set, we can obtain a realistic measure of their predictive capabilities and generalization to real-world scenarios. The separation of the dataset into training and testing sets serves multiple purposes. Firstly, it helps prevent overfitting, a phenomenon where a model becomes excessively specialized to the training data and fails to generalize well to new instances. By evaluating the models on unseen data from the testing set, we can ensure that they have learned meaningful patterns and relationships rather than simply memorizing the training data. Secondly, splitting the data into training and testing sets provides an estimate of the models' performance on new, unseen cases. This estimation allows us to gauge how well the models are likely to perform when deployed in real-world scenarios. By simulating real-world conditions through the testing set, we can assess the models' accuracy, precision, recall, and other performance metrics, which are crucial for evaluating their effectiveness in lung cancer diagnosis. Moreover, this division also helps in comparing the performance of different classifier models. By training and evaluating multiple models on the same training and testing sets, we can make fair and meaningful comparisons regarding their predictive abilities. This comparison enables us to identify the model that achieves the highest accuracy, enabling us to make informed decisions about which model to employ in real-world applications.

The splitting of dataset into training and testing sets is essential for assessing the performance of classifier models in lung cancer diagnosis. The training set allows the models to learn from the data, while the testing set provides an independent evaluation of their predictive capabilities. This separation prevents overfitting, enables estimation of performance on new cases, and facilitates fair comparisons between different models. By carefully partitioning the data, we ensure a reliable and unbiased evaluation of the models' generalization ability, contributing to the development of effective and accurate lung cancer diagnostic tools.

C. Model Building and Evaluation

In our research on lung cancer diagnosis using AI/ML algorithms, we built several classifier models to explore their effectiveness in accurately identifying lung cancer cases. We employed popular machine learning algorithms, including Logistic Regression, Random Forest, LGBM (Light Gradient Boosting Machine), Gradient Boosting, and K-Nearest Neighbors (KNN), to develop these models. Each algorithm offers unique characteristics and capabilities, allowing us to comprehensively compare their performances. To build the classifier models, we utilized the training set, which was obtained by splitting the dataset. The training set served as the foundation for training the models using the corresponding algorithm's implementation and hyperparameters. Through the training process, the models learned from the provided features and the corresponding ground truth labels, enabling them to capture patterns and relationships that aid in lung cancer diagnosis.

By leveraging Logistic Regression, we created a model

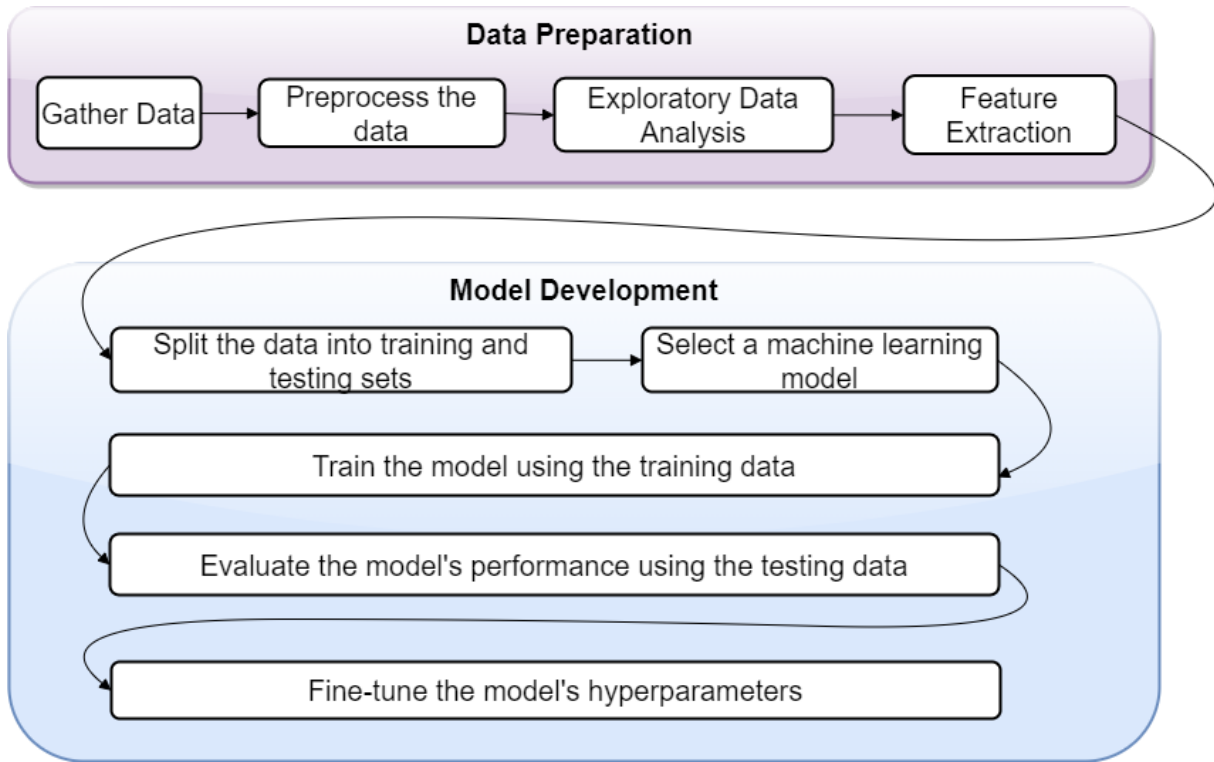


Fig. 4. Block representation of the proposed model.

TABLE II. DESCRIPTIVE STATISTICS OF FEATURES

	Age	Smoking	YF	Anexity	PP	CD	Fatigue	Allergy	Wheezing	AC	Coughing	SOB	SD	CP
count	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000	309.000
mean	62.673	1.563	1.570	1.498	1.502	1.505	1.673	1.557	1.557	1.557	1.579	1.641	1.469	1.556
std	8.210	0.497	0.496	0.501	0.501	0.501	0.470	0.498	0.498	0.498	0.494	0.481	0.500	0.497
min	21.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
25%	57.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
50%	62.000	2.000	2.000	1.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	1.000	2.000	2.000
75%	69.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
max	87.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000

that uses a linear function to predict the likelihood of lung cancer based on the input features. Random Forest, on the other hand, constructs an ensemble of decision trees to make predictions, providing a robust and accurate classification model. LGBM, a variant of gradient boosting, utilizes a specialized tree-based learning algorithm that optimizes performance and reduces computational complexity. Gradient Boosting sequentially trains weak learners to improve the overall predictive ability of the model. Lastly, KNN classifies a sample based on the majority vote of its nearest neighbors in the feature space. In evaluating the effectiveness of the classifier models, we employed key performance indicators, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of the predictions made by the models. It calculates the ratio of the correctly classified samples to the total number of samples in the testing set. Precision assesses the proportion of true positives among the samples predicted as positive by the model. Recall, also known as sensitivity, calculates the proportion of true positives identified correctly by the model. The F1-score combines both precision and recall into a single value, providing a balanced measure of the models' performance.

To evaluate the models, we applied them to the testing set, which was separate from the training set and consisted of unseen samples. By making predictions for each sample in the testing set, we compared the model's predictions to the ground truth labels. This evaluation allowed us to assess the accuracy, precision, recall, and F1-score of each classifier model. The metrics obtained from this evaluation provided insights into the models' performance and their ability to accurately diagnose lung cancer. Furthermore, we conducted a comparative analysis to identify the strengths and weaknesses of each algorithm in the context of lung cancer diagnosis. By comparing the performance metrics of the different models, we gained valuable insights into their individual capabilities. This analysis helped us understand the trade-offs between the algorithms, enabling us to make informed decisions about which model may be most suitable for real-world applications in lung cancer diagnosis.

Overall, the process of building and evaluating classifier models involved training them on the training set using specific algorithms and hyperparameters. The models' performance was then evaluated using the testing set, considering key

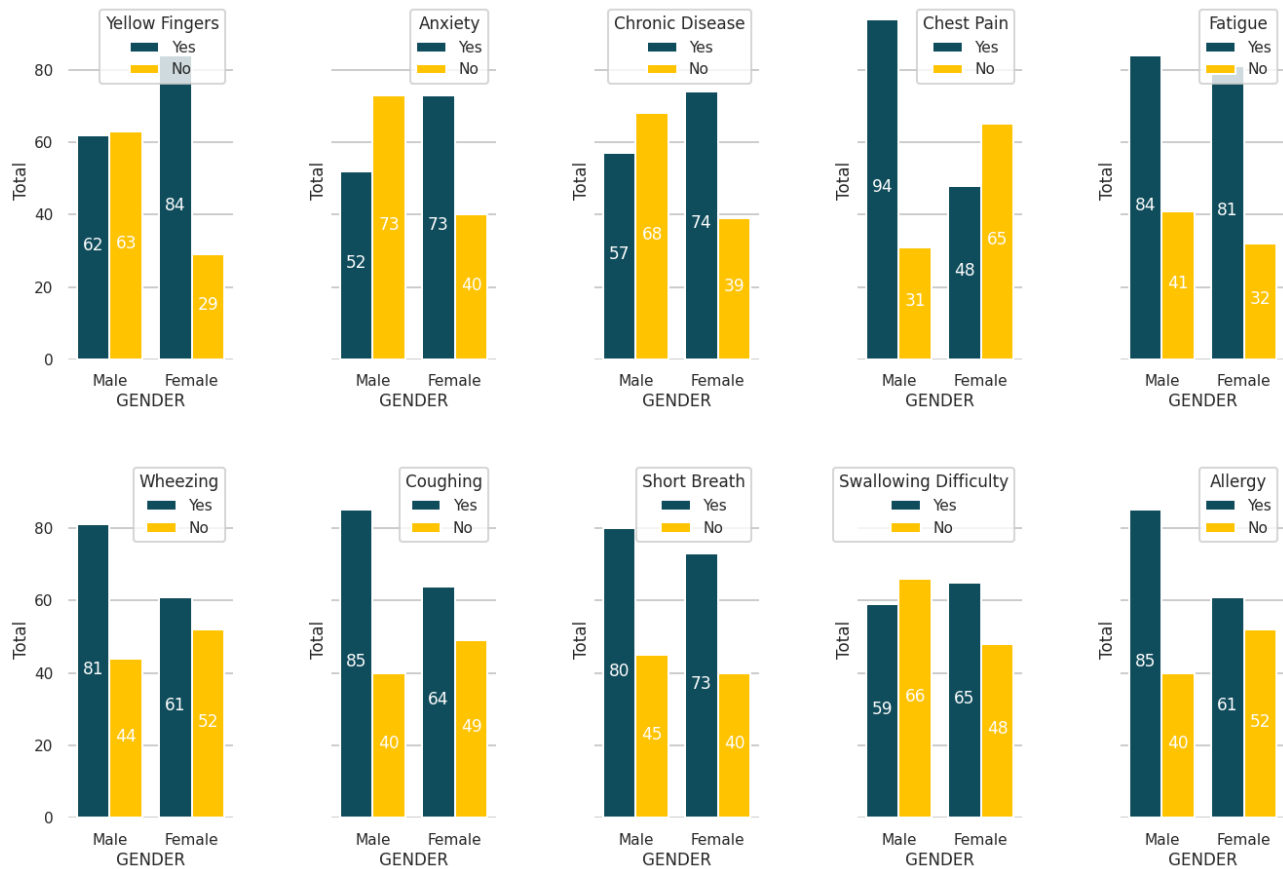


Fig. 5. Exploratory data analysis for lung cancer diagnosis.

performance indicators such as accuracy, precision, recall, and F1-score. Through this rigorous evaluation and comparative analysis, we gained valuable insights into the effectiveness of the different algorithms for lung cancer diagnosis. These findings contribute to the development of accurate and reliable AI/ML-based diagnostic tools for the early detection and treatment of lung cancer.

V. RESULTS AND DISCUSSION

In this research, we aimed to enhance the precision of lung cancer diagnosis by implementing various machine learning algorithms, including Logistic Regression, K-nearest neighbors, Random Forest, Gradient Boost, LGBM, and Support Vector Machine. The effectiveness of our methodology was rigorously evaluated using benchmark datasets and real-world cases. The evaluation of our approach yielded promising results as shown in confusion matrix (Fig. 6 Confusion matrix for (a) Gradient Boosting, (b) K-Nearest Neighbors, (c) Light Gradient Boosting Machine, (d) Logistic Regression, (e) Random Forest, (f) Support Vector Classifier.), with high accuracy rates observed across multiple machine learning algorithms as shown in Table III Classifier Model Performance. Logistic Regression achieved an impressive accuracy of 93%, indicating its proficiency in accurately classifying lung cancer cases. Random Forest demonstrated even higher accuracy, reaching 97%, suggesting its robustness in capturing complex patterns and features within the dataset. LGBM achieved an accuracy

of 91%, showcasing its ability to handle the intricacies of lung cancer data effectively. Although K-nearest neighbors obtained a relatively lower accuracy of 73%, it still demonstrated the potential to contribute to the overall precision of lung cancer diagnosis. These results underscore the potential of leveraging

TABLE III. CLASSIFIER MODEL PERFORMANCE

Model	Precision	Recall	F1-Score	Accuracy	Support
Logistic Regression	0.88	1.00	0.94	0.93	59
KNN	0.86	0.54	0.67	0.73	59
Random Forest	0.95	1.00	0.98	0.97	59
Gradient Boosting	0.90	0.47	0.62	0.71	59
LightGBM Classifier	0.94	0.86	0.90	0.91	59
SVM	0.50	1.00	0.67	0.50	59

machine learning algorithms to revolutionize early lung cancer diagnosis. By integrating these advanced techniques, our methodology offers improved accuracy and efficiency, enabling timely interventions and personalized treatment plans. Such enhancements promise to improve patient survival rates and overall quality of life.

Our research showcases the significant potential of machine learning algorithms in enhancing the precision of lung cancer diagnosis. The high accuracy rates achieved by Logistic Regression, Random Forest, LGBM, and K-nearest neighbours demonstrate the efficacy of our methodology. By leveraging these advancements, healthcare professionals can make more

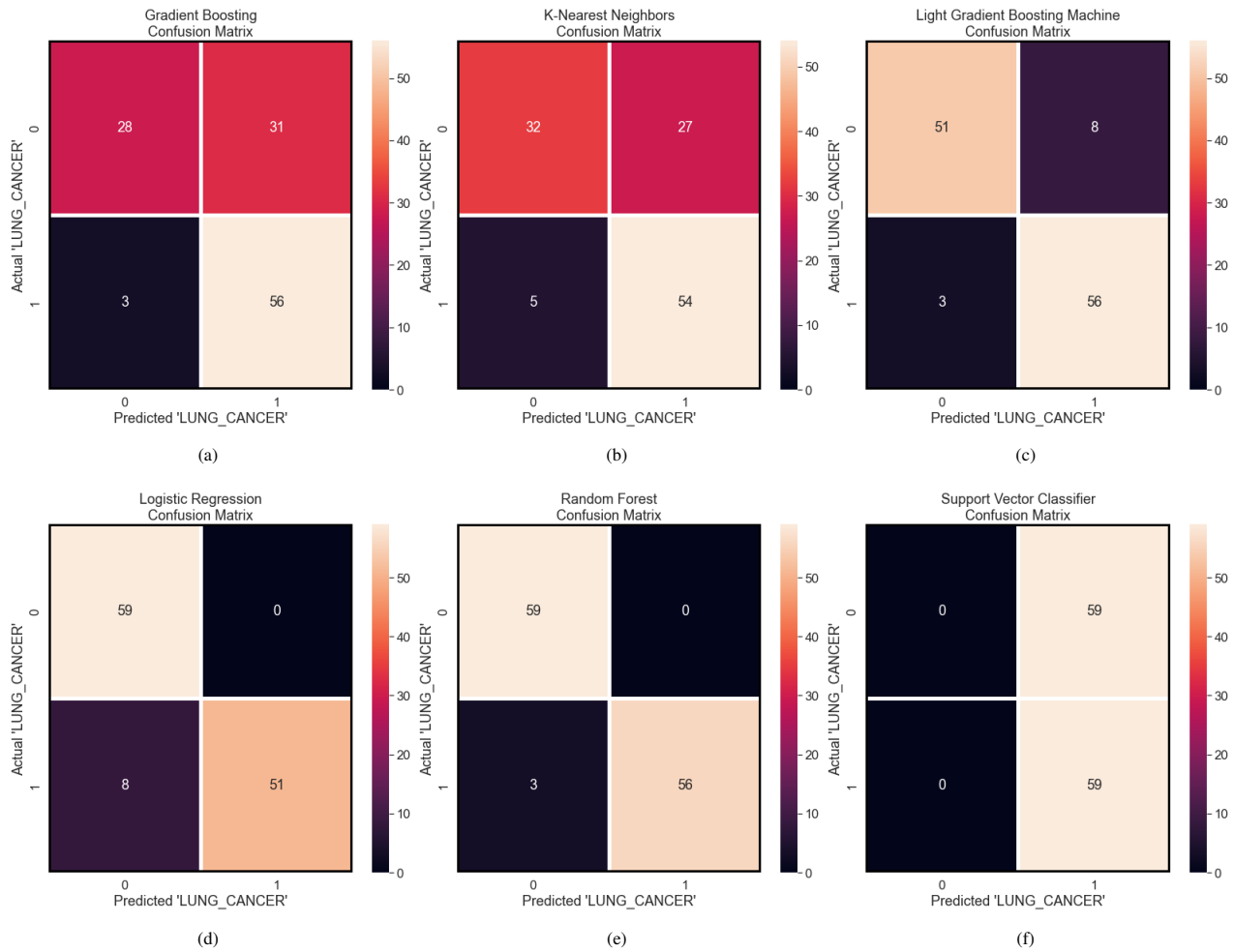


Fig. 6. Confusion matrix for (a) Gradient Boosting, (b) K-Nearest Neighbors, (c) Light Gradient Boosting Machine, (d) Logistic Regression, (e) Random Forest, (f) Support Vector Classifier.

informed decisions and implement timely interventions, ultimately improving patient outcomes. Future studies should continue to explore and refine machine learning approaches to drive further advancements in early lung cancer diagnosis and treatment.

Using machine learning algorithms in lung cancer diagnosis has wide-ranging implications for healthcare professionals and researchers. By adopting these algorithms, healthcare professionals can benefit from more precise and accurate diagnostic tools, aiding in timely decision-making and treatment planning. Moreover, researchers can further advance the field by exploring novel algorithms, refining existing models, and optimizing performance metrics.

VI. CONCLUSION

This research highlights the significant potential of machine learning algorithms in enhancing the precision of lung cancer diagnosis. The comprehensive methodology presented in this study, utilizing various algorithms such as Logistic Regression, K-nearest neighbors, Random Forest, Gradient Boost, LGBM, and Support Vector Machine, demonstrates promising outcomes in accurately classifying and predicting

lung cancer cases. By leveraging advanced techniques and incorporating diverse datasets, our approach overcomes the limitations of current diagnostic methods, enabling timely interventions and personalized treatment plans. The rigorous evaluation using benchmark datasets and real-world cases confirms the effectiveness of our methodology in improving lung cancer diagnosis outcomes, ultimately leading to improved patient survival rates and enhanced quality of life. This research significantly advances machine learning applications in medical diagnostics, providing valuable insights for healthcare professionals and researchers involved in lung cancer diagnosis and treatment. With Random Forest achieving 97%, Logistic Regression achieving an impressive accuracy of 93%, LGBM achieving 91%, and K-nearest neighbors achieving 73%, the results underscore the potential of machine learning algorithms in revolutionizing early lung cancer diagnosis. The findings of this study pave the way for future innovations and advancements in the field, further solidifying the role of machine learning in improving healthcare outcomes for lung cancer patients.

ACKNOWLEDGMENT

The authors acknowledge the support from REVA University for the facilities provided to carry out the research.

REFERENCES

- [1] Lewis, P. D., Lewis, K. E., Ghosal, R., Bayliss, S., Lloyd, A. J., Wills, J., ... & Mur, L. A. (2010). Evaluation of FTIR spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC cancer*, 10(1), 1-10.
- [2] Mathew, C. J., David, A. M., & Mathew, C. M. J. (2020). Artificial intelligence and its future potential in lung cancer screening. *EXCLI journal*, 19, 1552.
- [3] Makaju, S., Prasad, P. W. C., Alsadoon, A., Singh, A. K., & Elchouemi, A. (2018). Lung cancer detection using CT scan images. *Procedia Computer Science*, 125, 107-114.
- [4] Pradhan, K., & Chawla, P. (2020). Medical Internet of things using machine learning algorithms for lung cancer detection. *Journal of Management Analytics*, 7(4), 591-623.
- [5] Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Medicine*, 13(1), 1-17.
- [6] Nittas, V., Daniore, P., Landers, C., Gille, F., Amann, J., Hubbs, S., ... & Blasimme, A. (2023). Beyond high hopes: A scoping review of the 2019–2021 scientific discourse on machine learning in medical imaging. *PLOS Digital Health*, 2(1), e0000189.
- [7] Yang, X., Man, J., Chen, H., Zhang, T., Yin, X., He, Q., & Lu, M. (2021). Temporal trends of the lung cancer mortality attributable to smoking from 1990 to 2017: a global, regional and national analysis. *Lung Cancer*, 152, 49-57.
- [8] Kim, H. C., Kim, S. H., Kim, T. J., Kim, H. K., Moon, M. H., Beck, K. S., ... & Choi, C. M. (2022). Five-year overall survival and prognostic factors in patients with lung cancer: results from the Korean Association of Lung Cancer Registry (KALC-R) 2015. *Cancer Research and Treatment: Official Journal of Korean Cancer Association*, 55(1), 103-111.
- [9] Chaturvedi, P., Jhamb, A., Vanani, M., & Nemade, V. (2021, March). Prediction and classification of lung cancer using machine learning techniques. In *IOP conference series: materials science and engineering* (Vol. 1099, No. 1, p. 012059). IOP Publishing.
- [10] Hussain Ali, Y., Sabu Chooralil, V., Balasubramanian, K., Manyam, R. R., Kidambi Raju, S., T. Sadiq, A., & Farhan, A. K. (2023). Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer. *Bioengineering*, 10(3), 320.
- [11] Hamann, H. A., Ver Hoeve, E. S., Carter-Harris, L., Studts, J. L., & Ostroff, J. S. (2018). Multilevel opportunities to address lung cancer stigma across the cancer control continuum. *Journal of Thoracic Oncology*, 13(8), 1062-1075.
- [12] Thallam, C., Peruboyina, A., Raju, S. S. T., & Sampath, N. (2020, November). Early stage lung cancer prediction using various machine learning techniques. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1285-1292). IEEE.
- [13] Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6), 954-961.
- [14] Liao, F., Liang, M., Li, Z., Hu, X., & Song, S. (2019). Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. *IEEE transactions on neural networks and learning systems*, 30(11), 3484-3495.
- [15] Wang, Y., Fu, J., Wang, Z., Lv, Z., Fan, Z., & Lei, T. (2019). Screening key lncRNAs for human lung adenocarcinoma based on machine learning and weighted gene co-expression network analysis. *Cancer Biomarkers*, 25(4), 313-324.
- [16] Naik, A., & Edla, D. R. (2021). Lung nodule classification on computed tomography images using deep learning. *Wireless personal communications*, 116, 655-690.
- [17] Huang, X., Sun, W., Tseng, T. L. B., Li, C., & Qian, W. (2019). Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic CT scans using deep convolutional neural networks. *Computerized Medical Imaging and Graphics*, 74, 25-36.
- [18] Wang, H., Zhou, Z., Li, Y., Chen, Z., Lu, P., Wang, W., ... & Yu, L. (2017). Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images. *EJNMMI research*, 7, 1-11.
- [19] Hosny, A., Parmar, C., Coroller, T. P., Grossmann, P., Zeleznik, R., Kumar, A., ... & Aerts, H. J. (2018). Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. *PLoS medicine*, 15(11), e1002711.
- [20] Singal, G., Miller, P. G., Agarwala, V., Li, G., Kaushik, G., Backenroth, D., ... & Miller, V. A. (2019). Association of patient characteristics and tumor genomics with clinical outcomes among patients with non-small cell lung cancer using a clinicogenomic database. *Jama*, 321(14), 1391-1399.
- [21] Mehta, K. S. (2020). Enhanced Lung Nodule Malignancy Suspicion Classifier Using Biomarkers, Radiomics and Image Features (Doctoral dissertation, University of Maryland, Baltimore County).
- [22] Gürsoy Çoruh, A., Yenigün, B., Uzun, Ç., Kahya, Y., Büyükçeran, E. U., Elhan, A., ... & Kayı Cangır, A. (2021). A comparison of the fusion model of deep learning neural networks with human observation for lung nodule detection and classification. *The British Journal of Radiology*, 94(1123), 20210222.
- [23] Saleh, A. Y., Chin, C. K., Penshie, V., & Al-Absi, H. R. H. (2021). Lung cancer medical images classification using hybrid CNN-SVM. *International Journal of Advances in Intelligent Informatics*, 7(2), 151-162.
- [24] Lu, Y., Liang, H., Shi, S., & Fu, X. (2021, August). Lung cancer detection using a dilated CNN with VGG16. In *2021 4th International Conference on Signal Processing and Machine Learning* (pp. 45-51).
- [25] Gu, Y., Chi, J., Liu, J., Yang, L., Zhang, B., Yu, D., ... & Lu, X. (2021). A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. *Computers in biology and medicine*, 137, 104806.
- [26] Wu, R., & Huang, H. (2022, November). Multi-Scale Multi-View Model Based on Ensemble Attention for Benign-Malignant Lung Nodule Classification on Chest CT. In *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1-6). IEEE.
- [27] Huang, S., Yang, J., Shen, N., Xu, Q., & Zhao, Q. (2023, January). Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. In *Seminars in Cancer Biology*. Academic Press.
- [28] Huh, J. E., Lee, J. H., Hwang, E. J., & Park, C. M. (2023). Effects of Expert-Determined Reference Standards in Evaluating the Diagnostic Performance of a Deep Learning Model: A Malignant Lung Nodule Detection Task on Chest Radiographs. *Korean Journal of Radiology*, 24(2), 155.
- [29] Lv, W., Wang, Y., Zhou, C., Yuan, M., Pang, M., Fang, X., ... & Lu, G. (2021). Development and validation of a clinically applicable deep learning strategy (HONORS) for pulmonary nodule classification at CT: A retrospective multicentre study. *Lung Cancer*, 155, 78-86.
- [30] Bilal, A., Sun, G., Li, Y., Mazhar, S., & Latif, J. (2022). Lung nodules detection using grey wolf optimization by weighted filters and classification using CNN. *Journal of the Chinese Institute of Engineers*, 45(2), 175-186.
- [31] Liu, W., Liu, X., Luo, X., Wang, M., Han, G., Zhao, X., & Zhu, Z. (2023). A pyramid input augmented multi-scale CNN for GGO detection in 3D lung CT images. *Pattern Recognition*, 136, 109261.

Generating Nature-Resembling Tertiary Protein Structures with Advanced Generative Adversarial Networks (GANs)

Mena Nagy A. Khalaf¹, Taysir Hassan A Soliman², Sara Salah Mohamed³

Information System Department-Faculty of Computer and Information, Assiut University, Assiut, 71515, Egypt^{1,2}
Mathematics and Computer Science Department-Faculty of Science, New Valley University, New Valley, 72713, Egypt³

Abstract—In the field of molecular chemistry, the functions, interactions, and bonds between proteins depend on their tertiary structures. Proteins naturally exhibit dynamism under different physiological conditions, as they alter their tertiary structures to accommodate interactions with other molecular partners. Significant advancements in Generative Adversarial Networks (GANs) have been leveraged to generate tertiary structures closely mimicking the natural features of real proteins, including the backbone and local and distal characteristics. Our research has led to the development of stable model ROD-WGAN, which is capable of generating tertiary structures that closely resemble those found in nature. Four key contributions have been made to achieve this goal: (1) Utilizing Ratio Of Distribution (ROD) as a penalty function in the Wasserstein Generative Adversarial Networks (WGAN), (2) Developing a GAN network architecture that fertilizes the residual block in generator, (3) Increasing the length of the generated protein structures to 256 amino acids, and (4) Revealing consistent correlations through Structural Similarity Index Measure (SSIM) in protein structures with varying lengths. These model represent a significant step towards robust deep-generation models that can explore the highly diverse set of protein molecule structures that support various cellular activities. Moreover, they provide a valuable source of data augmentation for critical applications such as molecular structure prediction, inpainting, dynamics, and drug design. Data, code, and trained models are available at <https://github.com/mena01/Generating-Tertiary-Protein-Structures-Resembling-Nature-using-Advanced-WGAN>.

Keywords—Molecular structure; protein structure; protein modeling; tertiary structure; generative adversarial learning; deep learning; proteomic

I. INTRODUCTION

Molecular structures have been extensively researched over the past century due to their significant impact on our understanding of the human body and its functioning, both in normal and pathological states. This has facilitated the identification of the molecular basis of various diseases and facilitated the development of new strategies for their prevention and treatment [1]. In recent years, the pivotal role of bioinformatics models in the analysis of the molecular basis of diseases, including infectious diseases and cancers such as gallbladder cancer [2], lung cancer [3], colon cancer [4], [5], and prostate cancer [6], has been increasingly recognized.

The function and interactions of molecules largely depend on their structure. Therefore, predicting the structure of molecules can provide insights into their functions and has implications for a wide range of applications, including

drug design [7], molecule structure prediction [8], molecular inpainting [9], and molecular dynamics [10].

There are four different structures that proteins can have: primary structures [11], secondary structures [12], tertiary structures [13], and quaternary structures [14]. In biological laboratories, there are traditional methods that are used to determine these protein structures, such as X-ray crystallography [15], nuclear magnetic resonance (NMR) [16], and cryogenic electron microscopy (cryo-EM) [17]. However, these methods can be time-consuming and resource-intensive.

The gap between the number of known protein sequences and the number of discovered tertiary structures has increased exponentially and continues to grow [18]. According to the Protein Data Bank (PDB) [19], only around 180 thousand protein structures have been identified, compared to the approximately 207 million known protein sequences according to Uniport/TrEMBL [20]. As data scientists working in the field of protein structure prediction, our role is to generate tertiary structures of proteins that accurately mimic natural protein structures by capturing the natural protein structures' distribution.

CASP (Critical Assessment of protein Structure Prediction) [21] evaluates models that predict protein structures, and the recent introduction of Google's DeepMind AlphaFold v2 [22] has achieved the greatest performance in this area. It is important to note that proteins are naturally dynamic molecules [23] that can adopt different tertiary structures to modulate their interactions with different partners.

The dynamics of proteins have garnered significant attention lately, as evidenced by recent studies [24], [25], [26] that examine the balance motions between the spike glycoprotein (Receptor-Binding Domain (RBD) of the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2)) and the human Angiotensin-converting enzyme 2 (ACE2) receptor.

The spike glycoprotein is flexible and can transition between a closed and partially open structure, allowing it to bind to the ACE2 receptor and act as a viral entry point into human host cells.

Therefore, it is important to detect the diverse protein structures that proteins can access to regulate interactions with their molecular partners. Obtaining a broad view of the structure space is thus a vitally important research problem, and much work [26] has been focused on modeling proteins to capture this broad view of the protein structure space.

However, this is a challenging task, and most research [27] relies on existing protein structure data or restricted physical models [28] to guide search algorithms to the pertinent regions of the structure space that are otherwise too vast [29].

Early models used angles between bonds of atoms to simulate protein structures [9], [30] but more recent work has used GANs and long short-term memory networks to generate protein structures based on alpha-Carbon [30]. Despite the promising results obtained from these models, there is still much work to accurately simulate the diversity of protein structure.

In [9], the researchers used GANs with backbone angles in the representation of tertiary proteins, but they expanded the training dataset to include more proteins with various structures. However, it was observed that the generated protein structures exhibited distortion. As a result, the researchers replaced the backbone angles with distance matrices, which incorporated either the distances between each pair of Carbon Alpha (CA) atoms in the protein's main chain or the distances between every atom in the protein [31]. In the latter, the number of atoms increases, leading to larger distance matrices that can be difficult and time-consuming to train.

Recently, GAN networks have been employed to predict contact maps for protein structures [32], [33]. In this context, a contact map is a matrix in which the value of each element is 1 if two CA amino acids are in contact and 0 otherwise.

In [34], researchers trained their autoencoder (AC) on structures obtained from molecular dynamics simulations, such as computational platforms. In [35], the researchers used Rosetta as a platform for protein structure prediction to train the AC of Variational Autoencoder (VAE) [36]. In both cases, the researchers did not use experimental protein structures from the Protein Data Bank (PDB). However, in GAN models, it is preferable to use experimental structures from PDB rather than computational platforms.

In [10], the author used distance matrices of CA and produced nine models based on Vanilla GANs, which include Vanilla GAN, vanilla GAN + TTUR, Vanilla GAN + SpecNorm, Vanilla GAN + VBN, Vanilla GAN + TTUR + SpecNorm, Vanilla GAN + TTUR + VBN, Vanilla GAN + SpecNorm + VBN, Vanilla GAN + TTUR + SpecNorm + VBN, and WGAN.

The model achieved the highest accuracy was WGAN, denoted here as $WGAN_{Rahman}$, but it did not accurately capture the backbone and exhibited poor accuracy in both short-range and long-range structures. Furthermore, the generated distribution deviated significantly from the natural distribution, where the average peptide bond lengths of $WGAN_{Rahman}$ at 128 amino acids for backbone, short-range, and long-range structures were 7.5 Å, 11.66 Å, and 26.144 Å, respectively. In comparison, the natural average peptide bond lengths for backbone, short-range, and long-range structures are 3.78 Å, 7.79 Å, and 21.3 Å, respectively.

In this paper, our objective is to create models using WGAN [37] to generate tertiary protein structures that exhibit similar features to the natural protein structures in terms of their backbone, local, and distal protein structures. Additionally, we aim to ensure that the distribution of the generated

tertiary protein structures is comparable to that of the real tertiary protein structures.

We represented the tertiary structure using a CA distance matrix, as described in [9], [10]. Our models were trained using data from the PDB [19], which contains a diverse set of protein structures with varying amino acid lengths. We increased the amino acid length in our models to 256 aa. Additionally, we adjusted the WGAN gradient penalty by incorporating the ratio of distribution that achieved high accuracy within only 10 epochs. This contrasts with the best of the previous methods, where the $WGAN_{Rahman}$ model [19] was found to be unstable and achieved acceptable accuracy only after 50 epochs. To enhance stability, we utilized residual blocks in the generator network.

To summarise, the main contributions of our model ROD-WGAN, which make it different from the other models, are as follows:

- 1) Enhancing the WGAN gradient penalty by introducing the Ratio of Distribution (ROD) concept
- 2) Incorporating the convolution layers and the residual blocks in the Generator network to generate superior tertiary protein structures.
- 3) Increasing the length of the generated protein to 256 aa.
- 4) Our research reveals consistent correlations in protein structures through the application of Structural Similarity Index Measure (SSIM). These findings provide valuable insights into the inherent relationships within protein structures.

In the subsequent sections, this paper embarks on a comprehensive journey through the foundational elements of our study. The groundwork is established in Section II, where we present our proposed methodology and its key components. Progressing further, Section III meticulously details the refinement and preprocessing of our training dataset. Moving to Section IV, a thorough evaluation of our models takes place, wherein we compare them to state-of-the-art counterparts. Subsequent sections delve into the interpretation of experimental outcomes in Section V, while our contributions are summarized, and potential avenues for future research are suggested in Section VI, concluding this paper.

II. PROPOSED METHODOLOGY

The Generative Adversarial Network (GAN) [38] is a sophisticated architecture that has garnered attention from researchers across various fields, particularly in computer vision [39], [40], [41]. GAN has been employed to generate tertiary protein structures that mimic the real tertiary protein structure. In fact, this process is even more daunting than generating images due to the various constraints involved in the protein's structure, such as the backbone and short- and long-distance features. Previous GAN models have fallen short in capturing all three features of the tertiary protein structure with the same level of accuracy, and the discrepancy between the generated and the natural distributions was not close enough. The subsequent sections will briefly introduce the GAN model architecture and explain our model, ROD-WGAN.

A. GAN

GAN [38] consists of two neural networks that compete with each other: the Generator (G) and the Discriminator (D). The generator is responsible for generating fake proteins that simulate natural proteins and aims to deceive the discriminator, while the discriminator distinguishes between fake and real proteins.

As they compete against each other, each network tries to outperform the other. The balance between G and D leads to an optimal state in which their loss is equal to 0.5. Mathematically, assuming x represents the real data and z represents the latent vector or noise data, G is the generator that minimizes the function expressed in Eq. (1), and D is the discriminator that maximizes it.

$$\min_G \max_D GAN(G, D) = E_{x \sim p_r(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log 1 - D(G(z))] \quad (1)$$

Where p_r denotes the real data distribution, p_z denotes the model distribution, z is the input to the Generator and is randomly selected from some simple noise distribution.

The GAN network has encountered many problems, the most important of which are vanishing gradients and network instability. In [37], researchers proposed a WGAN network that uses Wasserstein distance to make the network more stable and faster, avoiding many of the issues faced with the GAN. The WGAN harnesses the 1-Lipschitz function, which guarantees the value is generated in a specific space and is enforced by the gradient penalty. It also replaces the name of the discriminator with the critic. The WGAN loss function is shown in Eq. (2) as follows:

$$L = \underbrace{E_{\tilde{x} \sim p_g} [D(x)] - E_{x \sim p_r} [D(x)]}_{originalcriticloss} + \underbrace{\lambda E_{\tilde{x} \sim p_{\tilde{x}}} [(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2]}_{gradientpenalty} \quad (2)$$

Where \tilde{x} is composed of real data x and fake data \tilde{x} , which is defined as $\tilde{x} = G(z)$, using the following equation:

$$\tilde{x} = \epsilon x + (1 - \epsilon)\tilde{x} \quad (3)$$

B. Ratio of Distribution (ROD)

According to our experimental findings, we found that the sum of the values of each distance matrix remains largely consistent across different proteins with the same number of amino acids. For example, as shown in Fig. 1, the sum of values of the distance matrix of different proteins with a length of 128 aa is 375000 Angstrom Å.

To compute ROD, some steps are required:

- 1) Calculate the mean sum of the natural proteins' distances matrices with the same length on all batches denoted as μ_r , (only performed once).
- 2) Calculate the mean sum of the distance matrices of generated proteins with the same length for each batch, denoted as μ_f (performed every time the fake data is generated).

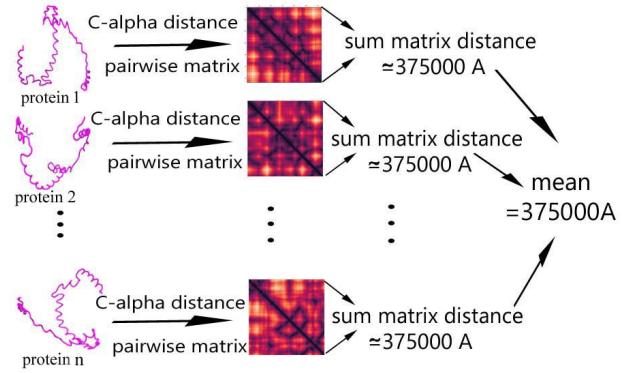


Fig. 1: Proteins with equal lengths of amino acids have equal sums of CA pairwise distance matrices.

- 3) Calculate the ratio of distribution ρ by dividing μ_r over μ_f .
- 4) Modify Equation (3) by adding the ratio of distribution ρ , as follows:

$$\rho = \frac{\mu_r}{\mu_f} \quad (4)$$

$$\hat{x} = \epsilon x + \rho * (1 - \epsilon)\tilde{x}$$

ROD ρ helped to generate close-to-real protein structures by capturing the backbone, short-range, and long-range features. In addition, the generated protein distribution is close enough to the real protein distribution, which accelerates and guarantees the stability of the learning process.

When μ_f is greater than μ_r , ρ is less than 1. Thus, we multiply the μ_f with the ρ to ensure that the mixed distance matrix value does not surpass the natural distance matrix value.

Conversely, when μ_f is smaller than μ_r , ρ is greater than 1. Thus, we multiply the μ_f with the ρ to ensure that the mixed distance matrix value does not fall below the natural distance matrix value.

In general, ρ controls the mixed distance matrix value to be aligned closely with the natural distance matrix value, as depicted in Fig. 2. The algorithm's steps are illustrated in Fig. 3.

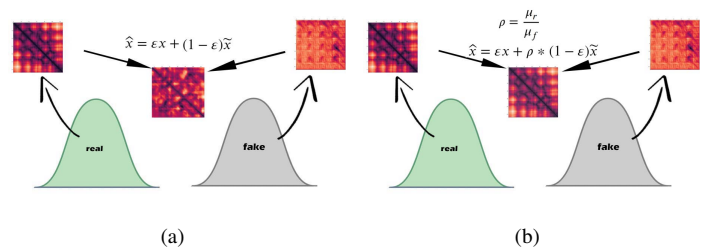


Fig. 2: The fake distribution trying to move to the real distribution by using a mixed distance matrix. a) Without the ratio of distribution and b) Using the ratio of distribution.

Algorithm 1 ROD-WGAN. We use default values of $\lambda=10$, $n_{critic}=5$, $\alpha=0.0001$, $\beta_1=0.05$, $\beta_2=0.999$.

Require: The gradient penalty coefficient λ , the number of critic iterations per generator iteration n_{critic} , the batch size m , Adam hyperparameters α, β_1, β_2 , and μ_r mean of sum natural protein distance matrix on batches. initial critic parameters w_0 , initial generator parameters θ_0

- 1: **while** θ has not converged **do**
- 2: **for** $i = 1, \dots, n_{critic}$ **do**
- 3: **for** $i = 1, \dots, m$ **do**
- 4: Sample real data $x \sim P_r$, latent variable $z \sim p(z)$, a random number $\sim \mathcal{U}[0, 1]$
- 5: $\tilde{x} \leftarrow G_\theta(z)$
- 6: $\mu_f \leftarrow \text{sum}(\tilde{x})$
- 7: $\rho \leftarrow \frac{\mu_r}{\mu_f}$
- 8: $\hat{x} \leftarrow \varepsilon x + \rho * (1 - \varepsilon)\tilde{x}$
- 9: $L^{(i)} \leftarrow D_w(\tilde{x}) - D_w(x) + \lambda(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2$
- 10: **end for**
- 11: $w \leftarrow \text{Adam}(\nabla_w \frac{1}{m} \sum_{i=1}^m L^{(i)}, w, \alpha, \beta_1, \beta_2)$
- 12: **end for**
- 13: Sample a batch of latent variables $Z_{i=1}^m \sim p(z)$
- 14: $\theta \leftarrow \text{Adam}(\nabla_\theta \frac{1}{m} \sum_{i=1}^m -D_w(G_\theta(z)), \theta, \alpha, \beta_1, \beta_2)$
- 15: **end while**

Fig. 3: The ROD-WGAN algorithm.

C. Model Architecture

The model architecture of GAN consists of two networks: the generator network and the discriminator network.

1) *The Generator Network Architecture:* The generator architecture of our models is illustrated in Fig. 4(a). The G network consists of convolution layers, which are fertilized by two residual blocks to enhance and accelerate the model’s learning. Table I provides the specifics of the generator network parameters for the 128 aa.

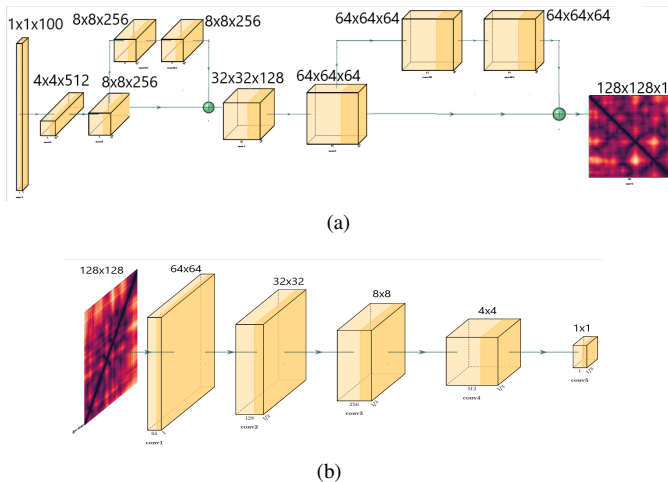


Fig. 4: The proposed architecture ROD WGAN a) Represents the Generator which consists of two Residual blocks and b) Represents the discriminator.

2) *The discriminator network architecture:* Fig. 4(b) shows the architecture of the discriminator. The discriminator takes the distance matrix of the protein structure produced by G and

TABLE I: THE LAYERS OF THE GENERATOR NETWORK ARCHITECTURE

layer		Details	filter	stride	padding
input		100*1*1	-	-	-
conv1		512*4*4	512*4*4	4	0
conv2		256*8*8	256*4*4	2	1
Residual1	conv	256*8*8	256*3*3	1	1
	conv	256*8*8	256*3*3	1	1
conv3		128*32*32	128*4*4	4	0
conv4		64*64*64	64*4*4	2	1
Residual2	conv	64*64*64	64*3*3	1	1
	conv	64*64*64	64*3*3	1	1
conv5		1*128*128	1*4*4	2	1

the distance matrix of the natural protein structure as input to differentiate the natural matrix from the generated one. The discriminator utilizes five convolution layers. Table II shows the discriminator network parameters for 128 aa .

TABLE II: LAYERS OF THE DISCRIMINATOR NETWORK ARCHITECTURE

layer		Details	filter	stride	padding
input		1*128*128	-	-	-
conv1		64*64*64	64*4*4	4	0
conv2		128*32*32	128*4*4	2	1
conv3		256*8*8	256*4*4	4	0
conv4		512*4*4	512*4*4	2	1
conv5		1*1*1	1*1*1	2	1

III. TRAINING DATASET

The dataset utilized comprised 115K protein structures sourced from PDB [19], with variations in their protein size. We calculated distance matrices that measured the distance between each pair of CA atoms within the protein’s main chain. As a result, the matrix distance size equaled $n*n$, where n equaled 64, 128, or 256 aa. It’s noteworthy that we are the first to create the 256aa structure; as the number of amino acids increases, the size of the distance matrix also increases, thereby increasing the complexity of model training.

To the extent of our knowledge, there has been no prior study on the generation of protein structures comprising 256 aa using the generative models. Our results align with recent studies on protein design via deep learning techniques [42], [43], as well as the latest advancements in the field of protein structure prediction and design [8], [9], [10].

IV. ASSESSMENT OF OUR MODELS

To evaluate the performance of our models as well as other state-of-the-art methods, we conducted several assessments, including i) Quantitative assessment, which involved evaluating the average peptide bond and comparing distributions; ii) Qualitative assessment; and iii) Convergence analysis.

A. Assessment on the Average Peptide Bond

The average peptide bond is calculated by summing all the entries along the main diagonal of the distance matrix and then dividing that sum by the length of the diagonal. To assess the quality of the generated protein tertiary structure, we compared its features (distance of backbone, short-range, and long-range) to those of natural proteins.

1) *Assessment on the backbone structure*: The backbone refers to the main diagonal of the generated distance matrix, which is constructed using every consecutive (i, i+1) CA pair where $0 < i < n-1$. In a natural protein, the ideal distance between two consecutive amino acids is 3.79\AA

2) *Assessment on the short-range structure and the long-range structure*: After computing the backbone, we can calculate the local structures by examining the short-range distance between consecutive (i, i+j) CA pairs, where j is between 1 and 4. In natural proteins, the ideal short-range distance is 7.8\AA . If we increase j beyond 4, we can determine the long-range distance, or distal structure. For 64 aa in a natural protein, the ideal long-range distance is 18.31\AA . While for 128 aa in a natural protein, the ideal long-range distance is 21.31\AA . Finally, when we computed it for 256 aa, we found a value of 25.01\AA , based on experimental data obtained from natural proteins.

B. Structural Similarity Index Measure

In our study and during our experiments, we made a noteworthy observation regarding the tertiary protein structures: there exists a consistent correlation between natural protein structures that have the same number of amino acids. When we calculate the SSIM between two different natural distance matrices, we obtained the following constant values for different lengths of distance matrices: 0.72 for distance matrices with a length of 64 aa, 0.69 for distance matrices with a length of 128 aa, and 0.68 for distance matrices with a length of 256 aa.

Based on these findings, we utilized SSIM as a loss function to enhance the similarity and correlation between the natural and the generated tertiary protein structures. We evaluated the SSIM score between the natural and the generated structures using Eq. (6).

$$SSIM_{(fake,real)} = \frac{2\mu_{real}\mu_{fake} + C_1}{\mu_{real}^2 + \mu_{fake}^2 + C_1} * \frac{2\sigma_{real}\sigma_{fake} + C_2}{\sigma_{real}^2 + \sigma_{fake}^2 + C_2} * \frac{2\sigma_{real*fake} + C_3}{\sigma_{real}\sigma_{fake} + C_3} \quad (5)$$

The formula includes several variables, such as the mean values of the real and the fake protein (μ_{real} and μ_{fake} , respectively), the standard deviation of the real and the fake protein (σ_{real} and σ_{fake} , respectively), as well as the cross-correlation ($\sigma_{real}\sigma_{fake}$) between the two proteins. Additionally, the formula contains three constants, labeled C_1 , C_2 , and C_3 equal to 0.01, 0.03, and 0.015, respectively [44].

The SSIM score ranges from 0 to 1, and a score closer to 1 indicates a greater level of correlation between the real and fake images, and vice versa. Therefore, we strive to achieve an SSIM score that is as close to natural as possible. Based on our experiments, if we calculate the SSIM between a natural distance matrix ($n*n$), where 'n' represents the length of the protein (either 64, 128, or 256), and itself, the SSIM value will always be one.

For example, if we take natural protein1 with a length of 64 aa, and natural protein2 with a length of 64 aa, and calculate

the SSIM between the distance matrices of these proteins, we will find the value to be 0.72. If we repeat the calculation for two different proteins, we will obtain the same constant value of 0.72. Hence, when calculating the SSIM between two different proteins with a length of 64 aa, the value is always constant at 0.72.

Similarly, if we take natural protein1 with a length of 128 aa, and natural protein2 with a length of 128 aa, and calculate the SSIM between the distance matrices of these proteins, we will find the value to be 0.69. If we repeat the calculation for different proteins, we will again obtain the same constant value of 0.69. Therefore, when calculating the SSIM between two different proteins with a length of 128 aa, the value is always constant at 0.69, regardless of the protein lengths.

Lastly, if we take natural protein1 with a length of 256 aa and natural protein2 with a length of 256 aa and calculate the SSIM between the distance matrices of these proteins, we will find the value to be 0.68. If we repeat the calculation for different proteins, we will once again obtain the same constant value of 0.68. Thus, when calculating the SSIM between two different proteins with a length of 256 aa, the value is always constant at 0.68.

C. Comparison of the Distribution

In GANs, we aim to capture the distribution of natural tertiary protein structures by approximating the generated distribution to the natural one. To measure the distance between the two distributions, we employ various metrics, such as the Earth Mover's Distance (EMD), Maximum Mean Discrepancy (MMD), and Bhattacharya Distance (BD).

1) *Earth Mover's Distance (EMD)*: The Earth Mover's Distance, also known as the Wasserstein distance [45], represents the minimum cost required to transform the generated distribution of tertiary protein structures to the natural distribution. EMD has been found to provide better perceptual dissimilarity than any other dissimilarity measure. EMD measures the distance between the two distributions, where a lower EMD value indicates higher similarity or proximity between the distributions, and a higher EMD value indicates lower similarity.

2) *Maximum Mean Discrepancy (MMD)*: Maximum Mean Discrepancy (MMD) [46] is a popular statistical test used to measure the distance between two distributions, $p(A)$ and $q(B)$. MMD is defined as the largest difference in the expectations of the mean of $A(\mu_A)$ and the mean of $B(\mu_B)$ over functions in the unit ball of a reproducing kernel Hilbert space (RKHS). MMD can be computed using Eq. (10). MMD measures the distance between the two distributions in the RKHS, where a lower MMD value indicates higher similarity or closeness between the distributions, and a higher MMD value indicates lower similarity.

$$MMD_{(A,B)} = \|\mu_A - \mu_B\|_H^2 \quad (6)$$

3) *Bhattacharya distance (BD)*: Bhattacharya Distance (BD) [47] is another measure of the distance between two distributions $p(a)$ and $q(a)$ on the same domain. BD can be computed by Eq. (11).

$$BD_{(p,q)} = -\ln(BC(p, q)) \quad (7)$$

where the Bhattacharaya Coefficient BC is

$$BC_{(p,q)} = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (8)$$

BC is an approximation that quantifies the degree of overlap between two samples drawn from distinct statistical distributions. A lower BD value indicates higher similarity or overlap between the two distributions, while a higher BD value indicates lower similarity.

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

We created ROD-WGAN model using the PyTorch framework on an RTX2080. We set the learning rate to 0.001 for both the critic and generator and used the Adam optimizer with b1 and b2 values of 0.5 and 0.999 respectively. The training time for one epoch of ROD-WGAN was approximately 17 minutes.

A. Quantitative Assessment

1) *The effect of ROD*: We have made significant progress in generating a distance matrix of tertiary protein structure using ROD. From the first ten epochs, we were able to capture the backbone, short-distance, and long-distance features of protein structures. As shown in Table III, our model, ROD-WGAN, outperformed WGAN without ROD and achieved better results that more closely resemble real protein structures. Furthermore, the distribution of the generated proteins is much closer to the natural protein distribution.

TABLE III: THE EFFECT OF ROD ON THE RESULTS OF BACKBONE, SHORT-RANGE, AND LONG-RANGE PROTEIN STRUCTURES ON JUST 10 EPOCH

Number of epoch	Features	Natural	WGAN without ROD	ROD-WGAN
10	Backbone	3.78	1.85	3.47
	Short	7.8	3.82	7.02
	Long	21.3	11.20	19.24

2) *Average peptide bond*: As mentioned earlier, we evaluated the distance matrix of the tertiary protein structure by considering the average length of peptide bonds in the backbone, short-range, and long-range distances. This method enabled us to accurately assess the similarity between the generated and natural distance matrices of the tertiary protein structure.

We assessed the quality of the backbone of the distance matrices generated by different models, namely ROD-WGAN, and $WGAN_{Rahman}$ [10]. As shown in Table IV and Fig. 5, we found that our model ROD-WGAN was able to capture the backbone, short-range, and long-range features of natural proteins more accurately than $WGAN_{Rahman}$ [10].

TABLE IV: DISTANCE FEATURES OF THE BACKBONE, THE SHORT-RANGE, AND THE LONG-RANGE FOR NATURAL AND GENERATED PROTEINS BY A VARIETY OF MODELS

		Natural	ROD-WGAN	$WGAN_{Rahman}$
64aa	Backbone	3.78	3.08	5.05
	Short	7.5	6.42	9.43
	Long	17.55	15.12	20.11
128aa	Backbone	3.78	3.014	7.506
	Short	7.8	6.58	11.66
	Long	21.31	19.24	26.144
256aa	Backbone	3.78	2.939	-
	Short	7.55	5.88	-
	Long	25.01	18.738	-

**The bold characters indicate the best evaluation scores.

TABLE V: SSIM BETWEEN THE NATURAL AND THE GENERATED DISTANCE MATRICES BY ROD-WGAN, AND $WGAN_{Rahman}$

SSIM	Natural	ROD-WGAN	$WGAN_{Rahman}$
64aa	72.47%	73.79%	72.02%
128aa	69.60%	70.19%	66.74%
256aa	68.13%	69.63%	-

3) *SSIM*: As previously mentioned, SSIM is a metric used to assess the quality and similarity between two distance matrices. Table V displays the performance of ROD-WGAN, and $WGAN_{Rahman}$ [10] on distance matrices of 64 aa, 128 aa, and 256 aa.

The ROD-WGAN model provided the highest protein structural similarity distance matrices, with ROD-WGAN being closer to the natural than $WGAN_{Rahman}$, particularly as the number of amino acids increased.

4) *Evaluation of the distribution distance*: We employed a variety of measurements, including EMD, MMD, and BD, to assess the disparity between the distribution of the generated distance matrices for the tertiary protein structure and the distribution of the natural distance matrices for the tertiary protein structure.

The line graph in Fig. 6 illustrates the performance of our models during the training process. Specifically, the plot depicts the changes in EMD, BD, and MMD values over time for each model. The consistently lower lines for our model, ROD-WGAN, as compared to the best state-of-the-art model $WGAN_{Rahman}$, serve as evidence of the accuracy of our models in generating protein structures that closely resemble those found in nature.

Fig. 6 illustrates that ROD-WGAN outperform $WGAN_{Rahman}$ [10] in the 64aa and 128aa regions. Furthermore, We observed that the MMD, BD, and EMD values obtained for the 256 aa region closely resemble those of the natural protein. This is noteworthy as the $WGAN_{Rahman}$ model [10] was originally implemented for a limited region of 128 aa and does not cover the entire 256 aa. Overall, the ROD-WGAN model accurately capture the distribution of the natural protein.

B. Qualitative Assessment

In Fig. 7, we present the 64*64, 128*128, and 256*256 distance matrices for the generated tertiary protein structures of ROD-WGAN, and $WGAN_{Rahman}$ models and the natural

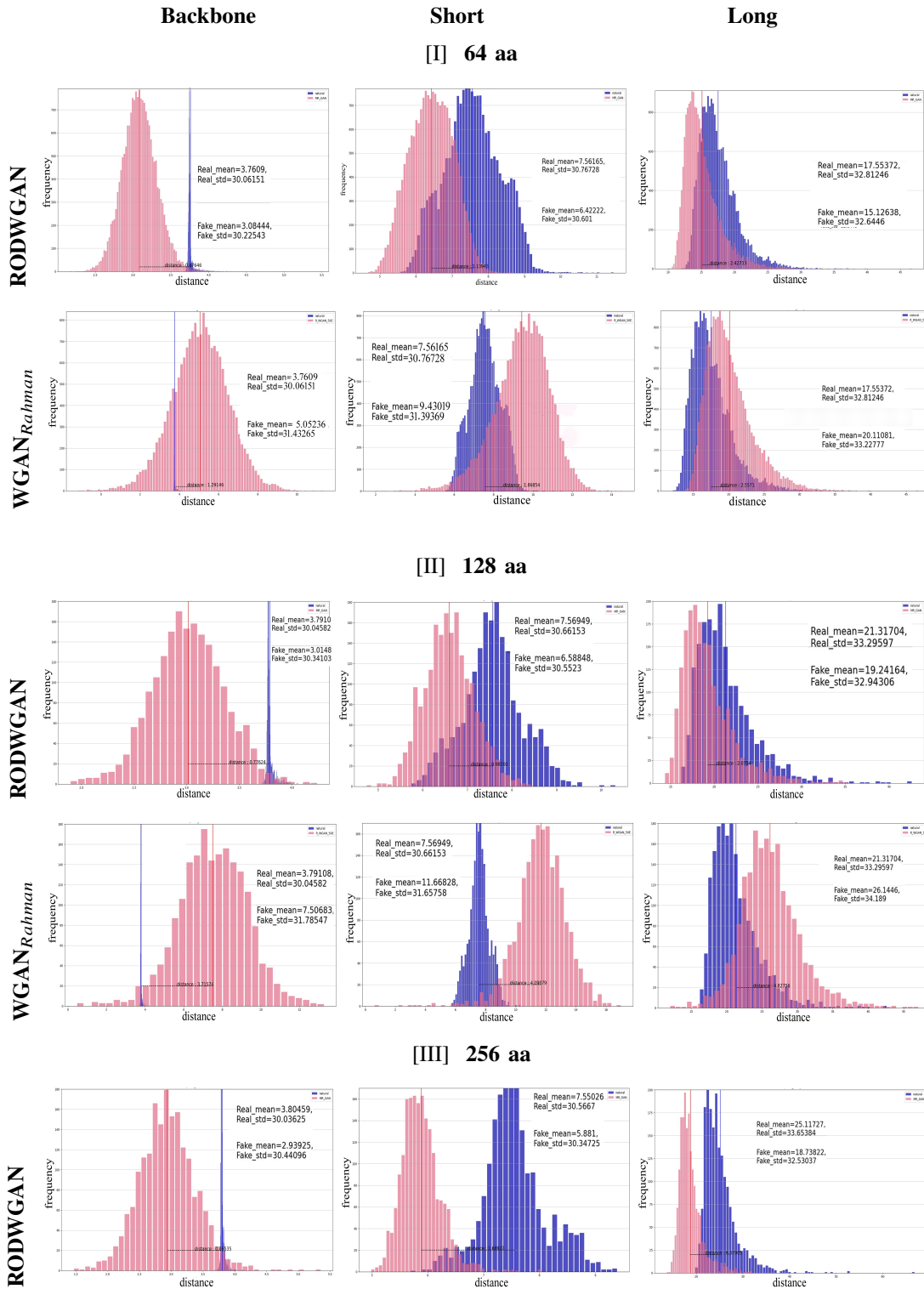


Fig. 5: The comparison has between the natural distribution(represented by the blue color)and the generated distribution(represented by the pink color). The *WGAN_Rahman* model was not implemented on 256 aa.

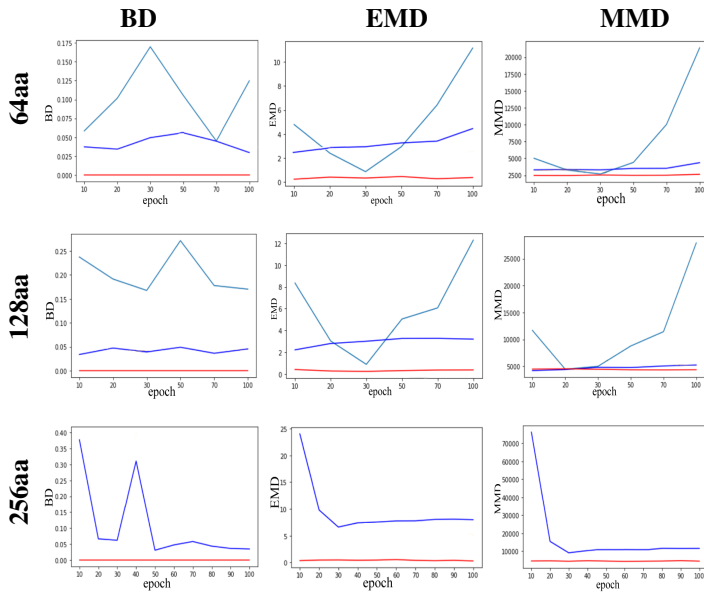


Fig. 6: Performance of ROD-WGAN, and $WGAN_{Rahman}$ on distributions for 64aa, 128aa, and 256aa.

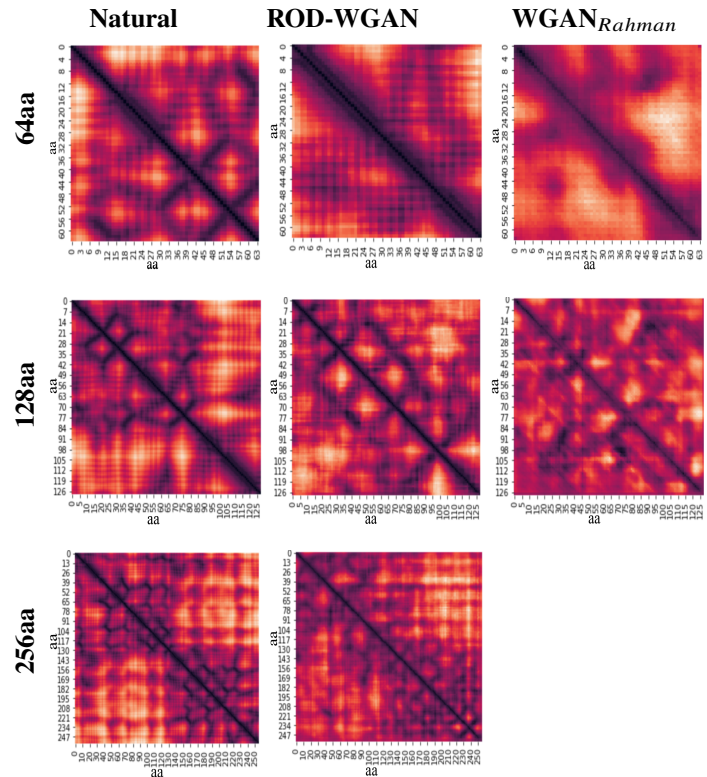


Fig. 7: Heatmaps visualized the Distance Matrices of the proteins' tertiary structures. The natural and generated distance matrices from various models ROD-WGAN, and $WGAN_{Rahman}$. The $WGAN_{Rahman}$ model [10] was not implemented on 256 aa.

structure. The heatmaps of these matrices were randomly selected from each model, with lighter colors indicating greater distance and darker colors indicating lower distance.

As seen in the 64aa and 128aa matrices, ROD-WGAN generated a clear heatmap distance matrix for the backbone, with a distinct dark diagonal, while the $WGAN_{Rahman}$ matrix was less clear. Furthermore, when we increase the length of the protein to 256aa, the ROD-WGAN model generate clear heat maps that have the backbone. To the best of our knowledge, there has been no previous report of generating protein structures with a length of 256 amino acids using the WGAN model. Our results are supported by recent surveys on protein design via deep learning [42], [43] and advances in protein structure prediction and design [8], [9], [10].

1) *Alternating Direction Method of Multipliers (ADMM)* : In our protein design study, we utilized the alternating direction method of multipliers (ADMM) [48] to convert the pairwise carbon alpha distance matrix (2d heatmap) to its equivalent 3d structure. We performed this for both the natural protein structures and those generated by various models ($WGAN_{Rahman}$, and ROD-WGAN). By employing the ADMM algorithm and implementing it with the software library [49], we were able to fold the distance matrices produced by our models to visualize the tertiary protein structures, as depicted in Fig. 8.

The visualization presented in Fig. 8 is crucial for evaluating the accuracy of the generated protein structures. It enables us to visually assess the overall shape of the generated structures and compare them against the natural structures. The ability to produce structures that closely resemble the natural structures is one of the most important characteristics of successful protein structure generation models. Therefore, the visualization in Fig. 8 provides an opportunity to validate the

performance of our models in generating protein structures' distance matrices. We observed that the structures generated from our model ROD-WGAN was much closer to the natural protein structures compared to those generated from the $WGAN_{Rahman}$ model.

C. Convergence Analysis in ROD-WGAN Model for Protein Structure Generation

In this study, we investigated the effectiveness of the generator (G) and discriminator (D) in reducing loss and achieving convergence during training epochs while ensuring the generation of high-quality protein structures. Our focus was on the ROD-WGAN model, designed specifically for protein structure generation. Fig. 9 illustrates the performance of the ROD-WGAN model on datasets comprising varying lengths of amino acids (aa), namely 64 aa, 128 aa, and 256 aa. Our objective was to assess the model's convergence capability and loss reduction across these datasets.

The results demonstrated that the ROD-WGAN model outperformed in reducing the overall generator loss. This indicates the significant improvement of the generator network (G) in generating realistic protein structures as the training progressed. Furthermore, the convergence between the total generator loss and the critic loss exhibited by the model

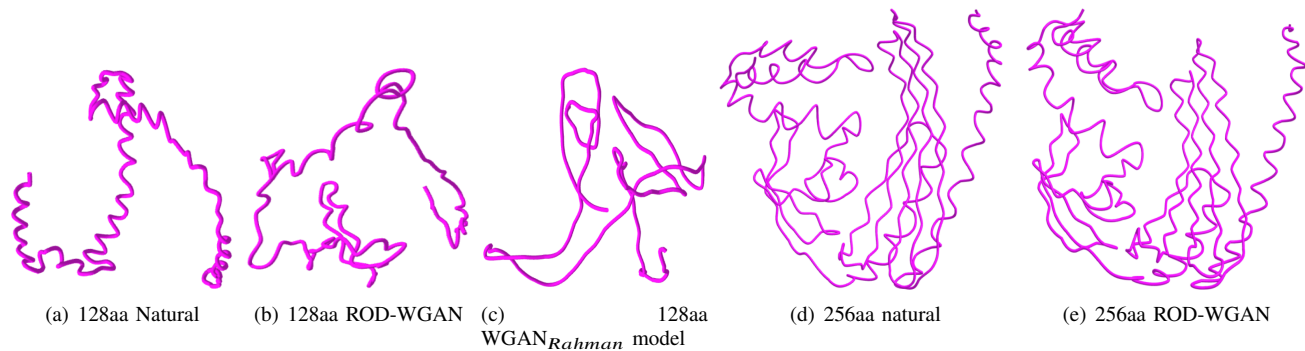


Fig. 8: The tertiary structure of protein structures a) The natural protein structure with a length of 128aa b) The structure of a protein generated from ROD-WGAN with a length of 128aa. c) The structure of a protein generated from WGAN_{Rahman} model [10] with a length of 128aa. d) The natural protein structure with a length of 256aa. e) The structure of a protein generated from ROD-WGAN with a length of 256aa.

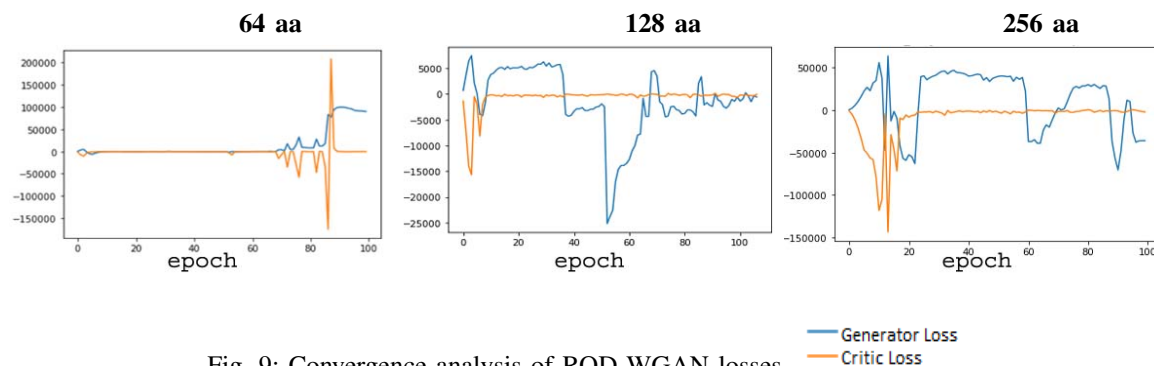


Fig. 9: Convergence analysis of ROD-WGAN losses.

indicated its stability and successful interplay between G and D.

VI. CONCLUSION AND FUTURE WORK

In this study, we not only focused on predicting the protein tertiary structure problem, but we were also interested in making the method of prediction more simple, more practical, and less laborious. Despite the success of AlphaFold in predicting the protein tertiary structure, there was still a need to search for another way that is easier, simpler, and does not require hundreds of TPUs (Tensor Processing Units).

We have developed models to generate distance matrices of proteins' tertiary structures in various amino acid lengths. Our proposed models are different from others in that they have the followings: 1) Modified the WGAN penalty equation by using the ROD 2) Developed Convolutional layers and enhanced it with the residual block 3) Applied on proteins with a length of 256 aa 4) our research uncovers consistent correlations in protein structures through the application of the SSIM. These findings provide valuable insights into the inherent relationships within protein structures, further enhancing the significance of our model.

In future work, we can try to generate tertiary protein structures based on distance and dihedral angle to increase the realism of the protein structures. We plan to work on

generating realistic and chemically accepting complex tertiary protein structures. We are also interested in tertiary protein structures as a data augmentation task for specific families of proteins that do not have an adequate amount of protein structures. We are also interested in the conditional GAN as a generative model by employing amino-acid sequences. Finally, we plan to build end-to-end models that start with a tertiary structure and end with different tertiary structures. They have formulated from its under various physiological conditions.

ACKNOWLEDGMENT

I acknowledge the help of the Data Science Lab Department of Information System, Faculty of Computers and Information, Assiut University, Egypt.

REFERENCES

- [1] S. Clayman and J. Heritage, *The news interview: Journalists and public figures on the air*. Cambridge University Press, 2002.
- [2] Y. Wang, A. Imran, A. Shami, A. A. Chaudhary, and S. Khan, "Decipher the helicobacter pylori protein targeting in the nucleus of host cell and their implications in gallbladder cancer: An insilico approach," *Journal of Cancer*, vol. 12, no. 23, p. 7214, 2021.
- [3] Y. Li, S. Khan, A. A. Chaudhary, H. A. Rudayni, A. Malik, and A. Shami, "Proteome-wide screening for the analysis of protein targeting of chlamydia pneumoniae in endoplasmic reticulum of host cells and their possible implication in lung cancer development," *Biocell*, vol. 46, no. 1, p. 87, 2022.

- [4] S. Khan, S. Zaidi, A. S. Alouffi, I. Hassan, A. Imran, and R. A. Khan, "Computational proteome-wide study for the prediction of escherichia coli protein targeting in host cell organelles and their implication in development of colon cancer," *ACS omega*, vol. 5, no. 13, pp. 7254–7261, 2020.
- [5] J. Li, M. Zakariah, A. Malik, M. S. Ola, R. Syed, A. A. Chaudhary, and S. Khan, "Analysis of salmonella typhimurium protein-targeting in the nucleus of host cells and the implications in colon cancer: an in-silico approach," *Infection and Drug Resistance*, vol. 13, p. 2433, 2020.
- [6] S. Khan, M. Zakariah, C. Rolfo, L. Robrecht, and S. Palaniappan, "Prediction of mycoplasma hominis proteins targeting in mitochondria and cytoplasm of host cells and their implication in prostate cancer etiology," *Oncotarget*, vol. 8, no. 19, p. 30830, 2017.
- [7] W. Yu and A. D. MacKerell, "Computer-aided drug design methods," in *Antibiotics*. Springer, 2017, pp. 85–106.
- [8] B. Kuhlman and P. Bradley, "Advances in protein structure prediction and design," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 11, pp. 681–697, 2019.
- [9] N. Anand and P. Huang, "Generative modeling for protein structures," *Advances in neural information processing systems*, vol. 31, 2018.
- [10] T. Rahman, Y. Du, L. Zhao, and A. Shehu, "Generative adversarial learning of protein tertiary structures," *Molecules*, vol. 26, no. 5, p. 1209, 2021.
- [11] C. I. Branden and J. Tooze, *Introduction to protein structure*. Garland Science, 2012.
- [12] M. Diener, J. Adamcik, A. Sánchez-Ferrer, F. Jaedig, L. Schefer, and R. Mezzenga, "Primary, secondary, tertiary and quaternary structure levels in linear polysaccharides: From random coil, to single helix to supramolecular assembly," *Biomacromolecules*, vol. 20, no. 4, pp. 1731–1739, 2019.
- [13] I. Rehman, C. C. Kerndt, and S. Botelho, "Biochemistry, tertiary protein structure," 2017.
- [14] S. Hauri, H. Khakzad, L. Happonen, J. Telemán, J. Malmström, and L. Malmström, "Rapid determination of quaternary protein structures in complex biological samples," *Nature Communications*, vol. 10, no. 1, p. 192, 2019.
- [15] M. Smyth and J. Martin, "x ray crystallography," *Molecular Pathology*, vol. 53, no. 1, p. 8, 2000.
- [16] K. Wüthrich, "The way to nmr structures of proteins," *Nature structural biology*, vol. 8, no. 11, pp. 923–925, 2001.
- [17] M. Carroni and H. R. Saibil, "Cryo electron microscopy to determine the structure of macromolecular complexes," *Methods*, vol. 95, pp. 78–85, 2016.
- [18] S. C. Pakhrin, B. Shrestha, B. Adhikari, D. B. Kc *et al.*, "Deep learning-based advances in protein structure prediction," *International Journal of Molecular Sciences*, vol. 22, no. 11, p. 5553, 2021.
- [19] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.
- [20] A. "Uniprot: the universal protein knowledgebase in 2021," *Nucleic acids research*, vol. 49, no. D1, pp. D480–D489, 2021.
- [21] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, and J. Moult, "Critical assessment of methods of protein structure prediction (casp)—round xiii," *Proteins: Structure, Function, and Bioinformatics*, vol. 87, no. 12, pp. 1011–1020, 2019.
- [22] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.
- [23] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *Nature chemical biology*, vol. 5, no. 11, pp. 789–796, 2009.
- [24] S. Majumder, D. Chaudhuri, J. Datta, and K. Giri, "Exploring the intrinsic dynamics of sars-cov-2, sars-cov and mers-cov spike glycoprotein through normal mode analysis using anisotropic network model," *Journal of Molecular Graphics and Modelling*, vol. 102, p. 107778, 2021.
- [25] R. Henderson, R. J. Edwards, K. Mansouri, K. Janowska, V. Stalls, S. Gobeil, M. Kopp, D. Li, R. Parks, A. L. Hsu *et al.*, "Controlling the sars-cov-2 spike glycoprotein conformation," *Nature structural & molecular biology*, vol. 27, no. 10, pp. 925–933, 2020.
- [26] H. Tian and P. Tao, "Deciphering the protein motion of s1 subunit in sars-cov-2 spike glycoprotein through integrated computational methods," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 17, pp. 6705–6712, 2021.
- [27] R. Clausen, B. Ma, R. Nussinov, and A. Shehu, "Mapping the conformation space of wildtype and mutant h-ras with a memetic, cellular, and multiscale evolutionary algorithm," *PLoS computational biology*, vol. 11, no. 9, p. e1004470, 2015.
- [28] E. Sapin, D. B. Carr, K. A. De Jong, and A. Shehu, "Computing energy landscape maps and structural excursions of proteins," *BMC genomics*, vol. 17, no. 4, pp. 433–456, 2016.
- [29] T. Maximova, E. Plaku, and A. Shehu, "Structure-guided protein transition modeling with a probabilistic roadmap algorithm," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 15, no. 6, pp. 1783–1796, 2016.
- [30] S. Sabban and M. Markovskiy, "Ramanet: Computational de novo protein design using a long short-term memory generative adversarial neural network," *BioRxiv*, p. 671552, 2019.
- [31] N. Anand, R. Eguchi, and P.-S. Huang, "Fully differentiable full-atom protein backbone generation," *ACM*, 2019.
- [32] H. Yang, M. Wang, Z. Yu, X.-M. Zhao, and A. Li, "Gancon: Protein contact map prediction with deep generative adversarial network," *IEEE Access*, vol. 8, pp. 80 899–80 907, 2020.
- [33] W. Ding and H. Gong, "Predicting the real-valued inter-residue distances for proteins," *Advanced Science*, vol. 7, no. 19, p. 2001314, 2020.
- [34] M. T. Degiacomi, "Coupling molecular dynamics and deep learning to mine protein conformational space," *Structure*, vol. 27, no. 6, pp. 1034–1040, 2019.
- [35] F. F. Alam and A. Shehu, "Variational autoencoders for protein structure prediction," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2020, pp. 1–10.
- [36] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.
- [37] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/892c3b1c6dcd52936e27cbd0ff683d6-Paper.pdf>
- [38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [39] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2439–2448.
- [40] C. Vondrick, H. Pirsiavash, and A. Torralba, "Generating videos with scene dynamics," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016. [Online]. Available: <https://proceedings.neurips.cc/paper/2016/file/04025959b191f8f9de3f924f0940515f-Paper.pdf>
- [41] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [42] Y. Ding, C. E. Lawrence, and A. E. Keating, "A deep learning framework for protein structure prediction," *Nature Methods*, vol. 19, no. 2, pp. 131–141, 2022.
- [43] S. Ovchinnikov and P. S. Huang, "Achieving high-resolution protein structure prediction with augmented neural networks," *Nature Communications*, vol. 12, no. 1, pp. 1–10, 2021.
- [44] J. Pessoa, "Pytorch_msssim: Pytorch implementation of ssim," <https://github.com/jorge-pessoa/pytorch-msssim>, 2021, accessed: April 8, 2023.

- [45] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International journal of computer vision*, vol. 40, no. 2, pp. 99–121, 2000.
- [46] A. Gretton, K. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, "A kernel method for the two-sample problem," *arXiv preprint arXiv:0805.2368*, 2008.
- [47] A. Mohammadi and K. N. Plataniotis, "Improper complex-valued bhattacharyya distance," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 5, pp. 1049–1064, 2015.
- [48] S. Ma, "Alternating direction method of multipliers for sparse principal component analysis," *Journal of the Operations Research Society of China*, vol. 1, no. 2, pp. 253–274, 2013.
- [49] K. You and X. Zhu, *ADMM: An R package for [brief description of the package]*, 2021, r package version 0.3.3. [Online]. Available: <https://CRAN.R-project.org/package=ADMM>

Prediction of Heart Disease using an Ensemble Learning Approach

Ghalia A. Alshehri, Hajar M. Alharbi
Department of Computer Science
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia

Abstract—The ability to predict diseases early is essential for improving healthcare quality and can assist patients in avoiding potentially dangerous health conditions before it is too late. Various machine learning techniques are used in the medical field. Nonetheless, machine learning is critical in determining the future of pharmaceuticals and patients' health. This is because the various classification techniques provide a high level of accuracy. However, because so much data are being gathered from patients, it becomes harder to find meaningful cardiac disease predictions. A vital research task is to identify these characteristics. Individual classification algorithms in this situation cannot generate flawless models capable of reliably predicting heart disease. As a result, higher performance might be achieved by using ensemble learning approaches (ELA), producing accurate cardiac disease predictions. In the present research work, we utilized an ELA for the early prediction of heart disease, using a new combination including four machine learning algorithms—adaptive boosting, support vector machine, decision tree, and random forest—to increase the accuracy of the prediction results. We used two wrapper methods for feature selection: forward selection and backward elimination. We used the proposed model with three datasets: the StatLog UCI dataset, the Z-Alizadeh Sani dataset, and the Cardiovascular Disease (CVD) dataset. We obtained the highest accuracy when using our proposed model with the Z-Alizadeh Sani dataset, where it was 0.91, while the StatLog UCI dataset was 0.83. The CVD dataset obtained the lowest accuracy, 0.73.

Keywords—Machine learning; ensemble learning; classification; disease prediction; heart disease

I. INTRODUCTION

Heart disease is a devastating illness that kills more people worldwide than other diseases. According to the annual statistical books of the Ministry of Health and the World Health Organization, heart disease caused 42% of deaths from non-communicable diseases in the Kingdom of Saudi Arabia (KSA) in 2010 [1]. Mortality from heart disease can be reduced if an accurate diagnosis is made early on. Modern medical science has demonstrated significant and effective ways of dealing with heart-related issues. Moreover, medical difficulties can now be addressed using artificial intelligence. Electrocardiogram (ECG), angiography screening, and blood tests are the most popular methods for detecting heart disease [2]. High cholesterol, blood pressure, and hypertension can all increase the risk of heart disease, but such signs may go unnoticed by the average person. Chest pain, breathlessness, and heart palpitations are frequent symptoms of heart disease. Angina, also known as angina pectoris, is a form of cardiac disease wherein the heart receives insufficient oxygen. Breathlessness

can occur due to heart failure when the heart becomes too weak to circulate blood. Some cardiac problems have no symptoms, particularly in the elderly and those with diabetes. When considering these factors, the healthcare industry must keep additional information about patients and their medications to generate diagnostic reports.

The advancement of computing and storage technology has allowed the healthcare industry to collect and retain routine medical data, allowing for more consistent and reliable support in medical choices. Patients' data are collected and maintained digitally in many developed countries. The information is then analyzed to make the required medical judgments regarding prediction, diagnosis, and treatment options [3]. Machine learning (ML) methods have been quite helpful in solving complicated classification and prediction problems [4]. One ML technique that can be used to predict future outcomes is classification. ML is crucial in recognizing cardiac illness from extensive data. ML aids in the decision-making process based on historical data. Classification, usually called supervised ML, predicts future events based on historical data. Medical ML employs techniques such as classification to generate insights and provide medical outcomes depending on the data [5]. In its most basic form, ML uses preprogrammed algorithms that learn and improve their operations by analyzing input data and making reasonable predictions. These algorithms tend to produce more accurate predictions as additional data are fed. Despite variations in classification, ML algorithms can be divided into three groups based on their objectives and how the underlying machine is trained: supervised, unsupervised, and semi-supervised [6]. A labeled training dataset trains the underlying algorithm in supervised ML techniques. The unlabeled test dataset is then assigned to the trained algorithm, categorizing it into similar categories [7]. It is feasible to gain insight into a patient's medical history and to provide clinical support through such an analysis. The risk that a person will develop heart disease can be predicted by training and testing classification algorithms. However, because the medical problem is so severe, the remedy necessitates greater classification accuracy, which is not provided by traditional classification algorithms.

Ensemble methods could be employed in this situation. More specifically, ensemble classification algorithms that integrate two or more classification techniques and generate the best prediction results are used to identify cardiac disease. Several ML techniques are used to treat heart illnesses due to their superior performance and capacity to comprehend

the relationships between features' input and output variables compared to experienced physicians or doctors. The values of different tests performed on a person are typically used as input features. Many classifications, clustering, and deep learning algorithms have been implemented by researchers worldwide. Despite this, given the rise in heart disease rates each year, newer ML methods should be implemented with significant features to improve the results of existing classification algorithms.

This research used an ELA containing four ML algorithms—adaptive boosting (AdaBoost), support vector machine (SVM), decision tree (DT), and random forest (RF)—to obtain the best results for predicting heart disease. We used two wrapper methods, forward selection, and backward elimination, for the feature selection step and analyzed unbalanced data.

II. BACKGROUND

A. Literature Review

We chose studies in the same field from 2017 to 2022, summarized them in Table I, and arranged them from oldest to newest. Yekkala et al. [3] studied using particle swarm optimization (PSO) and an ensemble classifier to predict cardiac disease. PSO was used as a feature selection method to eliminate the least-rated features, while ensemble methods were used to lower the misclassification rate and increase classification performance. The experiments showed that applying the bagged tree ensemble classifier to the PSO can significantly enhance learning accuracy. Dinh et al. [8] developed supervised ML models to detect individuals with cardiovascular, prediabetes, and diabetes diseases using the NHANES dataset [9]. Multiple ML models (logistic regression, SVMs, RF, and gradient boosting) were assessed for their classification performance and integrated into a weighted ensemble model to increase detection accuracy. David [5] used the StatLog dataset from the University of California–Irvine (UCI) data repository [10] to compare three algorithms—AdaBoost, Bagging, and Stacking—to identify the top ensemble classification method for predicting heart disease. According to one study, AdaBoost has been experimentally demonstrated to offer ideal results compared to its competitors. Liu et al. [11] suggested a unique ensemble learning method for medical diagnosis using imbalanced data. Using data preprocessing, training-based classifiers, and a final ensemble, they presented the SMOTE-CVCF integrated filter technique, C-SVM, and V-SVM with five kernel functions, a weighted fusion approach, and a SAGA method to optimize the weight vector. According to the empirical findings, the suggested ensemble learning method could outperform other cutting-edge categorization models. By randomly partitioning the dataset into smaller categories and employing a classification and regression tree (CART), Mienye et al. [12] improved an ML technique for forecasting the risk of heart disease. A modified version of the weighted aging classifier ensemble was employed to ensure the best performance, and a modified version of the weighted aging classifier ensemble was used to create a homogenous ensemble from several CART models. A novel coronary heart disease detection technique based on ML, such as classifier ensembles, was proposed by Tama et al. [13]. As a result, a two-tier ensemble was built, with certain ensemble classifiers serving as the foundation for another ensemble. The model

was evaluated using several heart disease datasets, and the proposed approach performed better than any base classifier in the ensemble. Yadav and Pa [14] proposed four algorithms for classifying data using trees and evaluated their accuracy, precision, and sensitivity. The M5P, random tree, reduced error pruning, and random forest ensemble approaches were employed in the first of the three experimental setups used for the analysis. The second experiment employed four tree-based techniques using recursive feature elimination, while Lasso regularization was used on top of the tree-based methods in the third trial. Predicting heart problems is just one of the many uses of this derivation process. Velusamy and Ramasamy [15] developed an ensemble algorithm with five features chosen based on feature importance, which was assessed using the Z-Alizadeh Sani dataset [16] and balanced using synthetic minority oversampling. When used on the balanced dataset, the weighted average voting (WAVEn) algorithm diagnosed coronary artery disease (CAD) with 100% accuracy, specificity, sensitivity, and precision. Tuncer et al. [17] proposed an ECG signal detection approach involving preprocessing, feature extraction, concatenation, selection, and classification. Fifteen sub-bands of ECG signals were generated during the preprocessing step. A maximum classification percentage of 96.60% was achieved for the MIT-BIH Arrhythmia dataset [18] using K-NN, and 97.80% accuracy was achieved using SVM for the St. Petersburg ECG dataset [19].

A summary of the literature review is shown in Table I. We abbreviated the labels of some metrics: Acc = accuracy, Sens = sensitivity, Spec = specificity, AUC = the area under the ROC curve, PPV = positive predictive value, NPV = negative predictive value, Prec = precision, MCC = Matthew's correlation coefficient, and Kappa = Cohen's kappa. As a reminder, a positive predictive value refers to precision, and recall refers to sensitivity. However, it is worth noting that the terminology may vary among different studies.

TABLE I: Summary of the Literature Review

Ref	Year	Method	Dataset	Best Result
[3]	2017	Ensemble methods (Bagged Tree, RF and AdaBoost) along with PSO	StatLog [20]	Bagged Tree Acc=100% Sens=100% Spec=100% PPV=100% NPV=100%
[8]	2019	A weighted ensemble model contains (Logistic Regression, SVM, RF, Gradient Boosting)	NHANES [9]	Prec=76% Recall=76% F1=76% AUC=83.9%

Continued on next page

TABLE I: Summary of the Literature Review (Continued)

[5]	2020	AdaBoost, Bagging and Stacking	StatLog [20]	AdaBoost Prec=81.2% Recall=80.6% F1=80.2%
[11]	2020	C-SVM and V-SVM with 5 kernel functions	UCI repository [21] + KEEL [22]	SPECTF heart dataset Prec=91.17% Recall=100% F1=95.38% AUC=95.98%
[12]	2020	Ensemble CART models	Cleveland [23] + Framingham [24]	Cleveland Acc=93% Framingham Acc=91%
[13]	2020	RF, Gradient Boosting, Extreme Gradient Boosting	Z-Alizadeh Sani [16] + StatLog [20], Cleveland [23], and Hungarian from UCI [10]	Z-Alizadeh Sani Acc=98.31% F1=96.60% AUC=98.70%
[14]	2020	M5P tree, Random tree and Error Reduced Pruning tree with RF Ensemble method	UCI repository [21]	Pearson Correlation feature selection on RF Acc=99.9% Sens=99.6% Spec=91.6%
[15]	2021	Heterogeneous ensemble method (K-NN, RF and SVM), with WAVEn	Z-Alizadeh Sani [16]	WAVEn method Acc=100% Kappa=100% Sens=100% Spec=100% Prec=100% F1=100% MCC=100%

Continued on next page

TABLE I: Summary of the Literature Review (Continued)

[17]	2022	LDA, K-NN, and SVM	MIT-BIH Arrhythmia [18] + St. Petersburg ECG [19]	MIT-BIH Arrhythmia with K-NN Acc=96.60% St.Petersburg ECG with SVM Acc=97.80%
------	------	--------------------	---	--

B. Dataset

We used three datasets for this study: the StatLog UCI dataset [10] and [20], the Z-Alizadeh Sani dataset [16], and the CVD dataset [25]. The StatLog dataset [20] from the UCI repository is commonly used for various cardiac illnesses. It contains 13 attributes and 270 cases. Information is included about the following attributes: Age; Sex; Chest pain type (Chp); Resting blood pressure (Bp); Serum cholesterol (Sch) in mg/dl; Fasting blood sugar (Fbs) greater than 120 mg/dL; Resting electrocardiographic result (Ecg); Maximum heart rate (Mhrt) achieved; Exercise induced angina (Exian); Old peak (Opk) = ST depression induced by exercise relative to rest; Slope of the peak exercise ST segment (Slope); Number of major vessels colored by fluoroscopy (Vessel); and Defect type (Thal). The target field “Class” indicates whether the patient has heart disease, with a value of 0 for no disease and 1 for disease.

The Z-Alizadeh Sani dataset [16] from the UCI repository contains 303 patient records, each with 54 features. The attributes are classified into four categories: (i) demographic, (ii) symptom and examination, (iii) ECG, and (iv) laboratory and echo features. Each patient falls into one of two categories: CAD or normal. If a patient’s diameter narrowing is greater than or equal to 50%, they are classified as having CAD; otherwise, they are classified as normal.

The CVD dataset [25] contains 70,000 patient records with the following different features: Age, Height, Weight, Gender, Systolic blood pressure (Ap_hi), Diastolic blood pressure (Ap_lo), Cholesterol, Glucose (Gluc), Smoking, Alcohol intake (Alco), and Physical activity (Active). The target class “Cardio” determines whether a patient is suffering from a cardiovascular illness (expressed as 1) or is healthy (shown as 0).

III. METHODOLOGY

A. Data Preprocessing

1) *Normalization*: In this step, we normalize the data. To enhance machine performance, an algorithm for learning data normalization is a preprocessing step that alters the attribute value in accordance with a standard scale or range. Examples of normalization methods include min-max, z-score, and decimal scaling [26]. There are many ML frameworks in the Python environment, such as sklearn [27]. This framework includes several helpful normalization algorithms, such as MinMaxScaler, MaxAbsScaler, StandardScaler, RobustScaler,

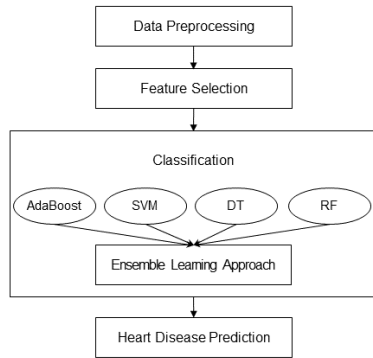


Fig. 1. Methodology framework.

and Normalizer. We used MinMaxScaler for normalization in this research.

2) *Imbalanced Data*: Imbalanced data are datasets with a highly uneven proportion of classes. Random Over Sampler, Random Under Sampler, Synthetic Minority Over-sampling Technique (SMOTE), and Tomek links are all examples of imbalanced data treatment [28]. To deal with the imbalanced data, we used SMOTE; the oversampling method involves producing synthetic instances rather than replacing oversampling for the minority class [29].

B. Data Splitting

When data is split, it is divided into two or more subsets. A two-part split is typically used to evaluate or test the data, and the other to train the model. Data splitting is a crucial feature of data science, especially for constructing data-collected models. This approach aids in the accuracy of data models and processes that employ data models, such as ML. We used cross-validation to split the dataset. Cross-validation is the most commonly used data-splitting approach in model selection. It separates the data into k distinct sections (k -folds) [30]. The validation set consists of one component (fold). The model is trained on the remaining $k-1$ portions (or folds), then applied to the validation set, and its prediction performance is recorded. This method was performed k times, resulting in each portion being utilized as a validation set just once. After averaging the recorded predicted performances, the optimal model parameter was selected with the best average predictive performance.

C. Feature Selection

Among practitioners, feature selection is a popular strategy for decreasing dimensionality. It seeks to choose a small subset of essential characteristics from the original collection based on specified criteria. Enhanced learning performance (e.g., increased learning accuracy for classification), reduced computation costs, and enhanced model interpretability are common outcomes of assessment criteria. Feature selection examples include filter, wrapper, and embedded methods [31]. Wrapper models assess the quality of features selected using a particular classifier and provide a simple and robust solution

to the feature selection problem independent of the learning machine used [32]. We used two wrapper methods: forward selection method and backward elimination. The forward selection method begins with no features. In each iteration, the feature that enhances the model performance is added until the model's performance is not improved by adding a new one. In contrast, the backward elimination method begins with the entire set of features and then gradually eliminates the least promising ones.

D. Classification

Classification is an ML approach to predicting data, such as group membership [33]. We used four classification models: AdaBoost, SVM, DT, and RF.

1) *AdaBoost*: One ensemble method for ML is called AdaBoost, or adaptive boosting. Decision trees of one level, or those with only one split [34], are AdaBoost's most frequently employed estimator. Decision stumps are another name for these trees.

2) *SVM*: SVM is one of the most renowned and practical techniques for dealing with data classifications, learning, and prediction challenges. The data points nearest the decision surface are support vectors [35]. It uses a hyperplane to classify data vectors in infinite dimensional space. The simplest type of SVM is the maximal margin classifier, which aids in determining the most basic classification problem of linearly separable training data with binary classification [36]. The maximal margin classifier determines the hyperplane with the most significant margin in real-world complexities. SVMs employ a variety of kernel methods. In this work, we used a linear kernel.

3) *DT*: The categorization process was improved by the straightforward DT modeling approach. All decision tree algorithms are typically built in two stages: (i) tree growth, where the training set is divided repeatedly based on local optimal criteria until the majority of the records in the partition have the same class label, and (ii) tree pruning, where the size of the tree is reduced to make it more comprehensible [35].

4) *RF*: A classification system using several decision trees is called the random forest approach. It uses bagging and feature randomization to create each tree, resulting in an uncorrelated forest of trees whose forecast by the committee is more accurate than any one tree [37].

E. Ensemble Learning Approach

Combining data fusion, data modeling, and data mining into a unified framework is the aim of ensemble learning. A set of features is first extracted from the ensemble learning data using various transformations [38]. Based on these learned attributes, a few learning algorithms produce mediocre predictions. Finally, utilizing voting systems, ensemble learning combines the valuable data from the quick findings to provide knowledge discovery and enhanced prediction performance.

F. Evaluations

One of the best ways to evaluate how well the proposed model performs is to examine its accuracy, PPV, NPV, sensitivity, specificity, AUC, MCC, and Kappa. Accuracy evaluates

how often the classifier guesses accurately [39]. The accuracy of a forecast can be defined as the ratio of correct predictions to total predictions, and is defined as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

The PPV reveals how many of the accurately anticipated cases were positive. Precision is functional when False positives are more of a worry than false negatives. The proportions of true positive and true negative results in diagnostic tests and statistics are PPV and NPV [40]. PPV and NPV describe the effectiveness of a diagnostic test or a similar statistical metric. A high result indicates that the statistic is accurate and can be determined, and is defined as

$$PPV(Precision) = \frac{TP}{TP+FP}$$

$$NPV = \frac{TN}{FN+TN}$$

According to Miao and Miao [41], sensitivity is the probability of successfully diagnosing the presence of cardiac disease in individuals, and is defined as

$$Sensitivity = Recall = \frac{TP}{TP+FN}$$

The probability of successfully identifying patients without cardiac disease is known as specificity, and is defined as

$$Specificity = \frac{TN}{FP+TN}$$

F1 provides a synthesis of the PPV and sensitivity measurements. It reaches its optimum when PPV and sensitivity are equal, and is defined as

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN}$$

The measure of a classifier’s ability to discriminate across classes is called the area under the curve (AUC) [40], and is defined as

$$AUC = Sensitivity - (1 - Specificity)$$

$$AUC = TPR - FPR$$

A statistical tool for assessing models is the MCC. It is accountable for determining the difference between anticipated and actual values, and is defined as

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$$

Kappa statistic is an excellent tool for handling difficulties involving multiple and unbalanced classes [42], and is defined as

$$Kappa = \frac{Po-Pe}{1-Pe}$$

IV. EXPERIMENTS AND RESULTS

A. Experiment 1

In this experiment, we used three datasets without applying the selection features. First, our proposed approach used the Statlog dataset. Using the Python environment AdaBoost classifier (with random state = 50), RF classifier (with n estimators = 600), SVM classifier (with linear kernel), and DT classifier (with random state = 500), we obtained the results mentioned in Table II.

In addition, we used the Z-Alizadeh Sani dataset with our proposed approach using the Python environment AdaBoost

TABLE II. EXPERIMENT 1 - RESULTS OF STATLOG UCI DATASET WITHOUT SELECTION

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.80	0.8	0.8	0.82	0.81	0.89	0.80	0.62	0.62
RF	0.84	0.82	0.82	0.87	0.83	0.92	0.84	0.69	0.70
SVM	0.82	0.8	0.8	0.84	0.81	0.91	0.81	0.65	0.66
DT	0.76	0.76	0.76	0.76	0.77	0.76	0.76	0.53	0.53
ELA	0.80	0.80	0.81	0.81	0.81	0.90	0.81	0.62	0.62

TABLE III. EXPERIMENT 1 - RESULTS OF Z-ALIZADEH SANI DATASET WITHOUT SELECTION

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.91	0.92	0.92	0.90	0.93	0.96	0.91	0.82	0.83
RF	0.92	0.92	0.91	0.91	0.92	0.97	0.91	0.84	0.86
SVM	0.87	0.89	0.89	0.85	0.89	0.93	0.87	0.74	0.74
DT	0.84	0.76	0.87	0.83	0.83	0.84	0.84	0.69	0.70
ELA	0.89	0.92	0.90	0.87	0.88	0.96	0.89	0.77	0.77

classifier (with random state = 400), RF classifier (with n estimators = 100), SVM classifier (with linear kernel), and DT classifier (with random state = 22); we obtained the results mentioned in Table III.

Furthermore, we used the CVD dataset without applying any selection feature methods using the Python environment AdaBoost classifier (with random state = 50), RF classifier (with n estimators = 500), SVM classifier (with linear kernel), and DT classifier (with random state = 200); we obtained the results mentioned in Table IV.

B. Experiment 2

In this experiment, we used the features obtained from three studies: Yekkala et al. [3], Velusamy and Ramasamy [15], and Chintan et al. [43]. The feature selection step was carried out by the three studies. In the first study, Yekkala et al. [3] used a PSO feature selection method to extract features from the Statlog dataset and extract seven features: Chp, Ecg, Mhrt, Exian, Opk, Vessel, and Thal. We took the same seven features they obtained and used them with the proposed model: the AdaBoost classifier (with random state = 1), RF classifier (with n estimators = 10), SVM classifier (with linear kernel), and DT classifier (with random state = 2), and obtained the results mentioned in Table V.

In the second study, Velusamy and Ramasamy [15] used feature selection from the Z-Alizadeh Sani dataset based on SVM. It is based on model information in which the model is trained to integrate the relationship between predictors for computing variable importance. They chose the top 12 features, namely: Atypical, Typical Chest pain, Age, Region with Regional wall motion abnormality (Region RWMA), Ejection Fraction (EF-TTE), Nonanginal Chest Pain (Nonanginal), Hypertension (HTN), FBS, Tinversion, BP, Diabetes Mellitus (DM), and TG. Then chose the following top five significant features: Typical Chest pain, Atypical, Age, Region RWMA, and EF-TTE. We used these features with the proposed model:

TABLE IV. EXPERIMENT 1 - RESULTS OF CVD DATASET WITHOUT SELECTION

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.72	0.63	0.81	0.77	0.56	0.79	0.70	0.45	0.46
RF	0.72	0.70	0.73	0.72	0.71	0.78	0.71	0.44	0.44
SVM	0.72	0.60	0.83	0.79	0.67	0.78	0.68	0.44	0.45
DT	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.28	0.28
ELA	0.68	0.69	0.67	0.68	0.68	0.77	0.68	0.37	0.37

TABLE V. EXPERIMENT 2 - RESULTS WITH 7 FEATURES FROM STATLOG UCI DATASET

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.77	0.74	0.74	0.80	0.76	0.85	0.76	0.55	0.56
RF	0.80	0.78	0.77	0.83	0.80	0.90	0.80	0.62	0.66
SVM	0.84	0.82	0.82	0.86	0.83	0.91	0.83	0.68	0.68
DT	0.76	0.80	0.80	0.76	0.80	0.76	0.77	0.53	0.55
ELA	0.80	0.81	0.82	0.82	0.83	0.91	0.80	0.59	0.61

TABLE VI. EXPERIMENT 2 - RESULTS WITH 12 FEATURES FROM Z-ALIZADEH SANI DATASET

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.89	0.89	0.89	0.89	0.91	0.96	0.88	0.78	0.79
RF	0.90	0.86	0.89	0.91	0.88	0.96	0.89	0.76	0.83
SVM	0.89	0.92	0.92	0.87	0.92	0.94	0.89	0.79	0.79
DT	0.85	0.87	0.87	0.83	0.88	0.84	0.85	0.69	0.70
ELA	0.91	0.93	0.91	0.90	0.92	0.96	0.90	0.82	0.81

the AdaBoost classifier (with random state = 1), RF classifier (with n estimators = 10), SVM classifier (with linear kernel), and DT classifier (with random state = 2). We obtained the results mentioned in Table VI and Table VII.

In the third study, Chintan et al. [43] used feature selection from the CVD dataset and estimated the mean arterial pressure (MAP) from the diastolic blood pressure (Ap_lo) and systolic blood pressure (Ap_hi) values for each instance. Patients' ages were initially given in days. Nonetheless, it was changed to years by dividing it by 365 to improve the analysis and prediction. They transformed the attributes of height and weight into body mass index (BMI), which may increase the performance of the heart disease prediction model. The nine features they obtained from the selection were Age, Gender, BMI, MAP, Cholesterol, Gluc, Smoke, Alco, and Active. We used these features with the proposed model: the AdaBoost classifier (with random state = 50), RF classifier (with n estimators = 500), SVM classifier (with linear kernel), and DT classifier (with random state = 200). We obtained the results mentioned in Table VIII.

C. Experiment 3

In this experiment, we used two wrapper methods for selecting features: forward selection and backward elimination. An iterative process called forward selection starts with the model having no features. The feature that best enhances our model is added in each iteration until the model's performance is not improved by adding a new variable. Backward elimination helps the model perform better by starting with

TABLE VII. EXPERIMENT 2 - RESULTS WITH 5 FEATURES FROM Z-ALIZADEH SANI DATASET

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.89	0.89	0.89	0.89	0.91	0.96	0.88	0.78	0.79
RF	0.90	0.86	0.89	0.91	0.88	0.96	0.89	0.76	0.83
SVM	0.83	0.87	0.87	0.81	0.86	0.92	0.84	0.67	0.67
DT	0.83	0.82	0.82	0.85	0.84	0.84	0.83	0.67	0.68
ELA	0.86	0.84	0.84	0.86	0.85	0.94	0.85	0.71	0.71

TABLE VIII. EXPERIMENT 2 - RESULTS WITH 9 FEATURES FROM CVD DATASET

Model	Acc	Sens	Spec	PPV	NPV	AUC	F1	Kappa	MCC
AdaBoost	0.72	0.64	0.64	0.76	0.69	0.78	0.69	0.44	0.44
RF	0.70	0.74	0.66	0.70	0.66	0.78	0.71	0.32	0.32
SVM	0.71	0.59	0.59	0.78	0.67	0.77	0.67	0.43	0.44
DT	0.63	0.60	0.60	0.63	0.62	0.63	0.62	0.26	0.26
ELA	0.66	0.65	0.67	0.67	0.66	0.74	0.66	0.33	0.33

all the features and removing the least important aspects one at a time. We keep doing this until we see no improvement when we remove features. We used four algorithms: AdaBoost, SVM, DT, and RF. We used R-squared as a measure of the performance of the feature-selection models. We applied it to the Statlog UCI, the Z-Alizadeh Sani, and the CVD datasets.

The result of feature selection with the Statlog UCI dataset is shown in Table IX, and it had the highest impact when using SVM with forward and backward methods. With the Statlog UCI dataset, we obtained the same result using both methods, but we chose the SVM with the forward method, as the number of significant features is less than 10. The results of the proposed model with the Statlog UCI dataset after using the SVM with the forward selection method are shown in Table XII. With the Z-Alizadeh Sani dataset, the highest R-squared result of the feature selection method was the SVM with the backward method using 26 features, as shown in Table X. The results of the proposed model with the Z-Alizadeh Sani dataset after using the SVM with the backward elimination method are shown in Table XIII.

With the CVD dataset, the result of feature selection is shown in Table XI, and it had the highest impact when using DT with backward, RF with forward, AdaBoost with forward, and AdaBoost with backward. We chose AdaBoost with the forward method. The results of the proposed model with the CVD dataset after using AdaBoost with the forward selection technique are shown in Table XIV.

We compared the results of the proposed model with previous studies that used one of the three datasets. When comparing the results of the proposed model with David's model [5], which used the Statlog UCI dataset, we found that the results of the proposed model outperformed their obtained results. In contrast, we obtained PPV = 0.83, which is higher than the [5] model (PPV = 0.812), and the proposed model got a higher sensitivity = 0.83, whereas the [5] model got a lower result (0.806 sensitivity). In addition, the [5] model had a lower F1 (F1 = 0.802) compared to the F1 of the developed model (F1 = 0.83). Unfortunately, David [5] was content with only three matrices to evaluate their results. It would have been better if they had used more matrices to comprehensively view the results. While Yekkala et al. [3] got 100% accuracy, sensitivity, specificity, PPV, and NPV. Some previous studies [13] and [15], which used the Z-Alizadeh Sani dataset obtained better results than the proposed model. Tama et al. [13] got an accuracy = 98.31%, F1 = 96.60% and AUC = 98.70%, while the proposed model obtained an accuracy = 91%, F1 = 89%, and AUC = 97%. Velusamy and Ramasamy [15] got 100% in accuracy, sensitivity, specificity, Kappa, PPV, F1, and MCC. However, this does not necessarily mean that their results are as good as in ML when 100% accuracy is achieved, which may indicate data overfitting.

TABLE IX: Experiment 3 - Best Wrapper Methods Results with Reduced Features Sets using Statlog UCI Dataset

Method	Number of features	Name of features	R2
Forward+DT	3	Exian, Vessel, Thal	0.34

Continued on next page

TABLE IX: Experiment 3 - Best Wrapper Methods Results with Reduced Features Sets using Statlog UCI Dataset (Continued)

Backward+DT	10	Age, Sex, Chp, Bp, Sch, Fbs, Ecg, Exian, Opk, Vessel	0.14
Forward+RF	3	Exian, Vessel, Thal	0.36
Backward+RF	12	Age, Sex, Chp, Bp, Sch, Fbs, Mhrt, Exian, Opk, Slope, Vessel, Thal	0.34
Forward+SVM	10	Age, Sex, Chp, Bp, Sch, Ecg, Mhrt, Exian, Vessel, Thal	0.44
Backward+SVM	12	Age, Sex, Chp, Bp, Sch, Fbs, Ecg, Mhrt, Exian, Opk, Vessel, Thal	0.44
Forward+AdaBoost	7	Sex, Chp, Fbs, Exian, Slope, Vessel, Thal	0.36
Backward+AdaBoost	8	Sex, Chp, Bp, Sch, Fbs, Opk, Vessel, Thal	0.33

TABLE X: Experiment 3 - Best Wrapper Methods Results with Reduced Features Sets using Z-Alizadeh Sani Dataset

Method	Number of features	Name of features	R2
Forward+DT	6	DM, EX-Smoker, Typical Chest Pain, EF-TTE, Region RWMA, VHD	0.42
Backward+DT	22	Age, Length, DM, HTN, Typical Chest Pain, Function Class, Q Wave, Tinversion, TG, LDL, ESR, K, Na, Region RWMA, Sex, CRF, CHF, DLP, Weak Peripheral Pulse, Dyspnea, LowTH Ang, LVH	0.31
Forward+RF	12	DM, EX-Smoker, Edema, Typical Chest Pain, St Elevation, Na, CHF, Region RWMA, Airway disease, Lung rates, LowTH Ang, Poor R Progression	0.42
Backward+RF	16	Age, HTN, BP, PR, Typical Chest Pain, St Elevation, Tinversion, CR, HDL, ESR, HB, Region RWMA, Obesity, CRF, Exertional CP, VHD	0.52
Forward+SVM	33	Age, Length, DM, HTN, Current Smoker, FH, PR, Typical Chest Pain, Q Wave, St Elevation, St Depression, Tinversion, FBS, CR, TG, HDL, BUN, HB, PLT, Lymph, Region RWMA, CRF, CVA, Thyroid Disease, CHF, DLP, Weak Peripheral Pulse, Lung rates, Dyspnea, Exertional CP, LowTH Ang, Poor R Progression, VHD	0.625

Continued on next page

TABLE X: Experiment 3 - Best Wrapper Methods Results with Reduced Features Sets using Z-Alizadeh Sani Dataset (Continued)

Backward+SVM	26	Age, Length, DM, HTN, Current Smoker, FH, PR, Typical Chest Pain, Function Class, Q Wave, Tinversion, CR, HDL, HB, K, WBC, Lymph, EF-TTE, Region RWMA, Sex, DLP, Airway disease, Lung rates, Dyspnea, Atypical, Nonanginal	0.626
Forward+AdaBoost	25	DM, HTN, Current Smoker, Edema, Typical Chest Pain, Q Wave, St Elevation, St Depression, CR, Region RWMA, Sex, CRF, CVA, CHF, Lung rates, Airway disease, Weak Peripheral Pulse, Systolic Murmur, Diastolic Murmur, Dyspnea, Exertional CP, LowTH Ang, Poor R Progression, VHD, LVH	0.50
Backward+AdaBoost	14	Age, BMI, DM, HTN, PR, ESR, Typical Chest Pain, CR, Tinversion, HB, WBC, EF-TTE, Region RWMA, Nonanginal	0.46

TABLE XI: EXPERMINT 3 - BEST WRAPPER METHODS RESULTS WITH REDUCED FEATURES SETS USING CVD DATASET

Method	Number of features	Name of features	R2
Forward+DT	8	Gender, Ap-hi, Ap-lo, Cholesterol, Gluc, Smoke, Alco, Active	-0.08
Backward+DT	7	Gender, Ap-hi, Cholesterol, Gluc, Smoke, Alco, Active	-0.07
Forward+RF	6	Ap-hi, Cholesterol, Gluc, Smoke, Alco, Active	-0.07
Backward+RF	8	Age, Height, Weight, Ap-hi, Ap-lo, Cholesterol, Gluc, Smoke	-0.16
Forward+SVM	5	Age, Height, Ap-hi, Cholesterol, Gluc	-0.08
Backward+SVM	10	Age, Gender, Height, Weight, Ap-hi, Cholesterol, Gluc, Smoke, Alco, Active	-0.08
Forward+AdaBoost	10	Age, Gender, Weight, Ap-hi, Ap-lo, Cholesterol, Gluc, Smoke, Alco, Active	-0.07
Backward+AdaBoost	10	Age, Gender, Weight, Ap-hi, Ap-lo, Cholesterol, Gluc, Smoke, Alco, Active	-0.07

V. CONCLUSION

The main goal of this study was to predict cardiac disease utilizing ELA, which included four ML algorithms: AdaBoost, SVM, DT, and RF. We applied it to three datasets: the StatLog UCI dataset, the Z-Alizadeh Sani dataset, and the CVD dataset.

TABLE XII. EXPERIMENT 3 - RESULTS OF PROPOSED METHOD WITH WRAPPER SELECTION USING STATLOG UCI DATASET

Table with 10 columns: Model, Acc, Sens, Spec, PPV, NPV, AUC, F1, Kappa, MCC. Rows include AdaBoost, RF, SVM, DT, ELA.

TABLE XIII. EXPERIMENT 3 - RESULTS OF PROPOSED METHOD WITH WRAPPER SELECTION USING Z-ALIZADEH SANI DATASET

Table with 10 columns: Model, Acc, Sens, Spec, PPV, NPV, AUC, F1, Kappa, MCC. Rows include AdaBoost, RF, SVM, DT, ELA.

We used two wrapper methods for the feature selection step, forward selection and backward elimination, and we dealt with the data imbalance using SMOTE. When using the proposed model with the StatLog UCI dataset, we obtained accuracy = 0.83, sensitivity = 0.83, specificity = 0.82, AUC = 0.90, PPV = 0.85, NPV = 0.83, F1 = 0.83, Kappa = 0.67, MCC = 0.68. When we used the Z-Alizadeh Sani dataset, we obtained accuracy = 0.91, sensitivity = 0.92, specificity = 0.90, AUC = 0.97, PPV = 0.93, NPV = 0.92, F1 = 0.89, Kappa = 0.81, MCC = 0.84. When using the CVD dataset, we obtained accuracy = 0.73, sensitivity = 0.63, specificity = 0.82, AUC = 0.77, PPV = 0.78, NPV = 0.69, F1 = 0.70, Kappa = 0.45, MCC = 0.46. In future work, we aim to collect a local dataset from King Abdullah Hospital - Bisha in the KSA, apply the proposed model to it, and improve the accuracy of the model. We will also use PSO and Gray Wolf Optimizer feature selection techniques that have shown promising results in disease prediction to further improve the model's performance.

TABLE XIV. EXPERIMENT 3 - RESULTS OF PROPOSED METHOD WITH WRAPPER SELECTION USING CVD DATASET

Table with 10 columns: Model, Acc, Sens, Spec, PPV, NPV, AUC, F1, Kappa, MCC. Rows include AdaBoost, RF, SVM, DT, ELA.

REFERENCES

[1] M. G. T. health, "Heart disease is the cause of 42% of deaths from non-communicable diseases in the kingdom," Oct 2013. [Online]. Available: https://www.moh.gov.sa/Ministry/MediaCenter/News/Pages/News-2013-10-30-002.aspx
[2] A. Rath, D. Mishra, G. Panda, and S. C. Satapathy, "Heart disease detection using deep learning methods from imbalanced ecg samples," Biomedical Signal Processing and Control, vol. 68, p. 102820, 2021.
[3] I. Yekkala, S. Dixit, and M. Jabbar, "Prediction of heart disease using ensemble learning and particle swarm optimization," in 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon). IEEE, 2017, pp. 691-698.
[4] G. Ranganathan, J. Chen, and A. Rocha, "Inventive communication and computational technologies," Ph.D. dissertation, Springer, 2021.
[5] H. B. F. David, "Impact of ensemble learning algorithms towards accurate heart disease prediction," ICTACT Journal on Soft Computing, vol. 10, no. 3, pp. 2084-2089, 2020.
[6] U. N. Dulhare, K. Ahmad, and K. A. B. Ahmad, Machine learning and big data: concepts, algorithms, tools and applications. John Wiley & sons, 2020.
[7] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," BMC medical informatics and decision making, vol. 19, no. 1, pp. 1-16, 2019.
[8] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," BMC medical informatics and decision making, vol. 19, no. 1, pp. 1-15, 2019.
[9] N. health and nutrition examination survey, "National health and nutrition examination survey," Dec. 2022, accessed: 2022-12-10.

[10] A. Janosi, W. Steinbrunn, M. Pfisterer, and R. Detrano, "Uci machine learning repository-heart disease data set," School Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 1988.
[11] N. Liu, X. Li, E. Qi, M. Xu, L. Li, and B. Gao, "A novel ensemble learning paradigm for medical diagnosis with imbalanced data," IEEE Access, vol. 8, pp. 171 263-171 280, 2020.
[12] I. D. Mienye, Y. Sun, and Z. Wang, "An improved ensemble learning approach for the prediction of heart disease risk," Informatics in Medicine Unlocked, vol. 20, p. 100402, 2020.
[13] B. A. Tama, S. Im, and S. Lee, "Improving an intelligent detection system for coronary heart disease using a two-tier classifier ensemble," BioMed Research International, vol. 2020, 2020.
[14] D. C. Yadav and S. Pal, "Prediction of heart disease using feature selection and random forest ensemble method," International Journal of Pharmaceutical Research, vol. 12, no. 4, pp. 56-66, 2020.
[15] D. Velusamy and K. Ramasamy, "Ensemble of heterogeneous classifiers for diagnosis and prediction of coronary artery disease with reduced feature subset," Computer Methods and Programs in Biomedicine, vol. 198, p. 105770, 2021.
[16] Z. Alizadehsani, R. Alizadehsani, and M. Roshanzamir, "Z-alizadeh sani data set," 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Z-Alizadeh+Sani
[17] T. Tuncer, S. Dogan, P. Plawiak, and A. Subasi, "A novel discrete wavelet-concatenated mesh tree and ternary chess pattern based ecg signal recognition method," Biomedical Signal Processing and Control, vol. 72, p. 103331, 2022.
[18] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, 2001.
[19] T. Viktor and A. Khaustov, "St.-petersburg institute of cardiological technics 12-lea d arrhythmia database," Circulation-Electronic, vol. 101, no. i23, pp. e215-e220, 2000.
[20] S. Sumbria, "Statlog (Heart) Data Set — kaggle.com," https://www.kaggle.com/datasets/shubamsumbria/statlog-heart-dataset, 2019, [Accessed 16-Jun-2023].
[21] A. Asuncion and D. Newman, "Uci machine learning repository," 2007.
[22] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework," Journal of Multiple-Valued Logic & Soft Computing, vol. 17, 2011.
[23] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease," The American journal of cardiology, vol. 64, no. 5, pp. 304-310, 1989.
[24] A. Bhardwaj, "Framingham heart study dataset," Apr 2022. [Online]. Available: https://www.kaggle.com/aasheesh200/framingham-heart-study-dataset
[25] S. Ulianova, "Cardiovascular Disease dataset — kaggle.com," https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset, 2019, [Accessed 16-Jun-2023].
[26] S. García, J. Luengo, and F. Herrera, Data preprocessing in data mining. Springer, 2015.
[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in python," the Journal of machine Learning research, vol. 12, pp. 2825-2830, 2011.
[28] R. Longadge and S. Dongre, "Class imbalance problem in data mining review," arXiv preprint arXiv:1305.1707, 2013.
[29] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.
[30] D. Krstajic, L. J. Buturovic, D. E. Leahy, and S. Thomas, "Cross-validation pitfalls when selecting and assessing regression and classification models," Journal of cheminformatics, vol. 6, pp. 1-15, 2014.
[31] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," Data classification: Algorithms and applications, p. 37, 2014.

- [32] N. Hasan and Y. Bao, "Comparing different feature selection algorithms for cardiovascular disease prediction," *Health and Technology*, vol. 11, pp. 49–62, 2021.
- [33] T. N. Phyu, "Survey of classification techniques in data mining," in *Proceedings of the international multiconference of engineers and computer scientists*, vol. 1. Citeseer, 2009, pp. 727–731.
- [34] R. E. Schapire, "Explaining adaboost," *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik*, pp. 37–52, 2013.
- [35] A. A. Soofi and A. Awan, "Classification techniques in machine learning: applications and issues," *J. Basic Appl. Sci.*, vol. 13, pp. 459–465, 2017.
- [36] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu *et al.*, "Top 10 algorithms in data mining," *Knowledge and information systems*, vol. 14, pp. 1–37, 2008.
- [37] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," *Expert systems with applications*, vol. 134, pp. 93–101, 2019.
- [38] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.
- [39] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, no. 4857, pp. 1285–1293, 1988.
- [40] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005. Proceedings 27*. Springer, 2005, pp. 345–359.
- [41] K. H. Miao, G. J. Miao *et al.*, "Mammographic diagnosis for breast cancer biopsy predictions using neural network classification model and receiver operating characteristic (roc) curve evaluation," *Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Bioinformatics*, vol. 3, no. 9, pp. 1–10, 2013.
- [42] D. Chicco, M. J. Warrens, and G. Jurman, "The matthews correlation coefficient (mcc) is more informative than cohen's kappa and brier score in binary classification assessment," *IEEE Access*, vol. 9, pp. 78 368–78 381, 2021.
- [43] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective heart disease prediction using machine learning techniques," *Algorithms*, vol. 16, no. 2, p. 88, 2023.

A Framework for Agriculture Plant Disease Prediction using Deep Learning Classifier

Mohammela Baljon*

Department of Computer Engineering,
College of Computer and Information Sciences,
Majmaah University, Majmaah, Saudi Arabia, 11952

Abstract—The agricultural industry in Saudi Arabia suffers from the effects of vegetable diseases in the Central Province. The primary causes of death documented in this analysis were 32 fungal diseases, two viral diseases, two physiological diseases, and one parasitic disease. Because early diagnosis of plant diseases may boost the productivity and quality of agricultural operations, tomatoes, Pepper and Onion were selected for the experiment. The primary goal is to fine-tune the hyperparameters of common Machine Learning classifiers and Deep Learning architectures in order to make precise diagnoses of plant diseases. The first stage makes use of common image processing methods using ml classifiers; the input picture is median filtered, contrast increased, and the background is removed using HSV color space segmentation. After shape, texture, and color features have been extracted using feature descriptors, hyperparameter-tuned machine learning (ML) classifiers such as k-nearest neighbor, logistic regression, support vector machine, and random forest are used to determine an outcome. Finally, the proposed Deep Learning Plant Disease Detection System (DLPDS) makes use of Tuned ML models. In the second stage, potential Convolutional Neural Network (CNN) designs were evaluated using the supplied input dataset and the SGD (Stochastic Gradient Descent) optimizer. In order to increase classification accuracy, the best Convolutional Neural Network (CNN) model is fine-tuned using several optimizers. It is concluded that MCNN (Modified Convolutional Neural Network) achieved 99.5% classification accuracy and an F1 score of 1.00 for Pepper disease in the first phase module. Enhanced GoogleNet using the Adam optimizer achieved a classification accuracy of 99.5% and an F1 score of 0.997 for Pepper illnesses, which is much higher than previous models. Thus, proposed work may adapt this suggested strategy to different crops to identify and diagnose illnesses more effectively.

Keywords—Suggested agricultural plant disease prediction system; tuned ML models; machine learning classifiers; plant disease detection; deep learning architectures

I. INTRODUCTION

The vast boundaries of the Kingdom of Saudi Arabia include an area of over two million square kilometres, or more than 80% of the whole landmass of the Arabian Peninsula. The country's position helps to explain its mild winters and hot, dry summers. It is between 15.2 and 32.6degrees north latitude and 34.1 and 55.5degrees [1] east longitude. Outside of the south-western highlands, where it rains more often in the summer, air receives less than 100 millimetres of precipitation annually. The natural springs in the Hue region provide the vast bulk of the water utilised in local farms in order to sustain water supply

and replenish aquifers, dams have been built at various locations around the country.

The agriculture industry in Saudi Arabia has received significant attention as part of the country's five-year growth goals. These initiatives aim to diversify the economy away from oil exports so that more people can eat and the quality of life can be kept high despite the population boom. The full potential of the country's agriculture sector is now being tapped upon. The quantity of land under cultivation is skyrocketed from around 435 thousand ha in 1980 to more than 1.5 million ha in 1990, [2] and a large part of that increase may be attributed to government encouragement and support. Two further agricultural academies emerged at the same time as new plant varieties were developed, the greenhouse business was launched, and massive agricultural undertakings were undertaken. In addition to meeting domestic need, the country currently exports food products such as wheat, dates, melons, poultry, fresh eggs, and milk [3]. Fig. 1 shows Saudi Arabia's market production.

Wheat, sorghum, barley, and millet are examples of important cereal crops; tomatoes and watermelons are examples of important vegetable crops; date palms, citrus trees, and grapevines are examples of important fruit crop species. The importance of fodder crops like lucerne is also significant. More over 1,100,000 acres, or around 81% of the total cultivated land area, is devoted to these crops. In 1990, wheat was grown on an estimated total of around 744 422 acres [4], or almost 55% of the total cultivated area. About 3.5 million metric tonnes were harvested from the crop.

Saudi Arabia Fruits and Vegetables Market: Export Volume in Metric Ton, Dates, 2017-2021

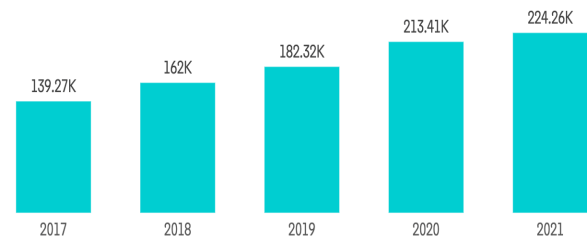


Fig. 1. Saudi Arabia vegetable market production metric.

* m.Baljon@mu.edu.sa

"Smart farming," often known as "precision agriculture," refers to the practise of using computer-based intelligent technologies in agriculture. Research in its infancy includes, but is not limited to, studies of intelligent irrigation systems, automated pesticide management, and the detection of plant diseases. Systematic methods of disease detection are crucial for the early diagnosis and prevention of certain conditions. Horticulturalists have an 88% success rate [5] in diagnosing plant diseases using standard diagnostic approaches. Despite this, it's a complicated procedure that calls for experience in the field. It may be difficult to undertake an inquiry of this sort in certain cases due to the crop's location. Several studies have proposed using Deep Learning Architectures (VGG 16, VGG 19, ResNet 50) or conventional Image Processing methods (Image filtering, contest stretching, segmentation, feature extraction, and disease classification) to classify various plant diseases. Traditional image processing techniques excel in situations when a large number of training data is unavailable. This means that several, fine-grained techniques, all fine-tuned to a high degree of precision, are required for correct sickness classification.

The structure of the paper is as follows; Section II includes study of existing methodology, Section III includes proposed methodology, Section IV includes Experimental analysis, Section V includes conclusion and future work.

II. LITERATURE SURVEY

A farmer's first priority should be the detection and eradication of plant diseases. The leaves are the first part of the plant to show damage from an approaching attack, and the changes that occur in the leaves are readily apparent. However, if you try to segment a picture of a sick leaf, you'll probably wind up with a jumbled, poorly lit composite. Because of this, we'll be taking a look at and ranking the latest classification algorithms for identifying problems with tomato leaves.

Effective convolutional neural network (CNN) was the name given to the deep CNN model developed by the team working on this project [2]. After the model was built and trained, it was put to the test by identifying images of healthy and diseased tomato leaves. The U-net model and a modified U-net model were put to the test and compared with regards to how effectively they segmented leaves. Experiments using six- and ten-class classification models were also undertaken in addition to the binary-classification ones. When it comes to segmenting images of leaves, only the U-net model was able to achieve an accuracy of 98.66 per cent. In contrast, EfficientNet-B7 showed stable performance across a range of classification tasks, from binary to six-class classifications. Using segmented pictures, the average accuracy of binary classification is 99.95%, while the average accuracy of six-class classification is 99.12%.

Using an image segmentation approach optimised for super pixels, the authors of [3] created a system for automatically recognizing and classifying tomato illnesses. A color-balance technique was employed during preproduction so that the optimal threshold for each kind of image collection could be identified. To further distinguish the leaves from the background, a ground-breaking method based on a histogram of gradients and color changes was used. A pyramid of the

histogram of gradients, a shape descriptor, and a grey level co-occurrence matrix (GLCM) are all components of a feature extraction approach that has been shown to be effective in distinguishing various medical presentations that are otherwise identical. The use of classifiers is pervasive in this investigation. Still, the best results towards the goal of the suggested framework were achieved by the random forest classifier, trained on a dataset of one hundred trees. When comparing the results of this study with those of many others that are quite similar to it [4,5,6,7], it was found that comparative analysis based on estimate parameters was the most accurate technique.

To classify potential tomato leaf diseases, the authors of [8] proposed a multi-class feature extraction strategy. The system is based on a deep CNN model that makes use of the attention method and the residual block. The findings show that the model is effective at picking up commonalities across diseases. Moreover, it uses the widely-used Plant Village dataset to get better results than a substantial body of prior deep learning literature. The overall positive identification percentage for this inquiry was 99.24%.

In [9], the feasibility of separating apart the various tomato diseases is examined. To boost model performance with little computational overhead, a lightweight CNN approach was developed and shown. The computational complexity, performance, and network architecture are only few of the areas that were explored for this essay. The used dataset consists of information from one healthy person and nine patients with various diseases. According to the results, if a model that is easy to understand and computationally efficient is created, the classification accuracy might be improved. Many different CNN architectures were developed for the tomato leaf disease detection study presented in [10]. LeNet, VGGNet, ResNet50, and Xception are only few of these designs. Scientists used 14,903 images of tomato plant leaves from the Plant Village dataset to build a deep Convolutional Neural Network (CNN). Both healthy and diseased plant leaves were shown in these photographs. The data showed that among all the tested designs, the fine-tuned VGGNet de-sign provides the best classification (99.25% accuracy) and achieves the lowest loss. This is true despite the fact that there is a significant time commitment associated with training and substantial financial outlay for the necessary technology

Grape plant leaf diseases are no match for the deep transfer learning-based model developed by the authors of [11]. In order to separate out the most crucial features, the authors built a fully connected layer. After that, redundant data was removed from the feature extractor vector using the variance method. With the use of images from the Plant Village dataset, the Efficient Net B7 deep architecture was retrained in this case. Then, logistic regression was utilized to further refine the collected data. By using this method, we were able to improve our categorization accuracy to 99.7%.

Using a dataset of 3,000 images of tomato leaves using the Google Collaborative Net-work (CN) model, the authors of [12] were able to accurately identify and classify nine different diseases and one healthy leaf class. Images are first preprocessed, then regions are separated, and finally the

findings are presented in this novel approach. The images are then processed further, which entails tuning the CNN model's hyper-parameters. The input picture is then analysed by the CNN, which pulls out features like as colors, borders, and textures. Prediction accuracy for the built categorization model is 98.49%.

According to the study published in [13], a CNN model with three convolutional layers, one max-pooling layer, and a filter count that can be adjusted between one and ten was developed. The authors of the Plant Village dataset used augmentation techniques to rectify the imbalance they discovered between the number of photos in each class. This model has an average accuracy of 91.2% and needs nearly 1.5 MB of storage space, whereas the pretrained model needs 100 MB.

The authors of [14] employed a transfer learning technique to construct a deep learning model that needed less training data, less computational resources, and less time to train. The scientists employed five different deep network topologies—MobileNet, Resnet50, Xception, Densenet121 Xception, and Shuffle Net—to extract characteristics. The authors experimented with many educational methods and tempos. With a classification accuracy of 97.10%, the DenseNet Xception easily triumphed over the competition.

According to [15], a deep residual network is built to identify tomato leaf diseases. To improve the remaining thick network, the authors chose to alter its structure. The model might be easily modified into a classification model with a 95% accuracy rate. The authors of [16] provide a system that can automatically detect and classify leaf diseases. To lessen the load on the computer's resources, the input photographs must be downscaled before the background can be removed and the photos separated. In order to extract features, the researchers in this work used two distinct deep learning models: VGG19 and Alex Net. Then, an ECOC-based SVM classifier was used to determine the identities of these characteristics. The results showed that VGG19 and Alex Net both achieved 98.8% and 98.9% classification accuracy, respectively.

In order to categorize the wide variety of diseases that might afflict soybean plants, the authors of [17] use multilayer perceptron deep learning and support vector machine techniques. In all, 19 diseases were correctly classified by the SVM. There were 683 examples in the dataset used; 643 were correctly classified while the remaining 40 had incorrect labels, for a classification accuracy rate of 94.14 per cent.

As was seen in the prior section of this article, the classification accuracies of various deep learning and machine learning algorithms created and contested for the detection of tomato leaf diseases vary widely. In addition, a few techniques have been created. In Table I, we can see a comparison of the algorithms utilized the diseases that may be identified, and the accuracy of the various classification systems that have been covered thus far.

TABLE I. A COMPARISON BETWEEN THE RECENTLY DEVELOPED CROP CLASSIFICATION SYSTEM

[2]	CNN model	3 disease classes	RF: 97%
[3]	Segmentation-based CNN	8 disease classes	99.23%
[4]	Light weight CNN	8 disease classes	99.25%
[5]	VGGNet- CNN	8 disease classes	VGGNet: 99.25%
[6]	Hybrid CNN (Hy-CNN)		
[7]	RNN Model	8 disease classes	97%
[8]	SVM model	8 disease classes	98.39%
[9]	ResNet-50 + SeNet	8 disease classes	97.8%
[15]	Restructured residual dense network	8 disease classes	95%
[16]	VGG19, Alex Net		91.2%
[17]	Deep learning, SVM		Densenet121 Xception

III. PROPOSED MATERIALS AND METHODS

A. Proposed First Phase DLPDS

The sequential method is used for the first phase of layer creation. This is due to the fact that building the model in layers is easiest using the sequential approach. To implement Softmax's input shape of (28, 28, 1), we utilise the add () function with the parameters conv2D, kernel _size, activation 'Elu,' and model layers. These are the fundamental building blocks of a convolutional neural network.

This procedure, which starts with the input picture and continues through the Conv2D layers and the flatten layers [18] connects the convolution and the other dense layers. Machine learning (ML) algorithms have become widely employed as artificial intelligence (AI) has advanced because ML succeeded in achieving emerges and cost-effective solutions to exploration of harvest yield [19]. The best result, which might be any of the values 0 through nine, is selected by setting the node in output to ten.

Each value is given a probability according to the Softmax activation, and the highest probability value is used as the forecast to determine whether the model needs more training to improve.

The agriculture sector makes use of validity measures to assess the performance of the disease segmentation technique for leaf blight. Blight, which affects tomatoes, is caused by the proteobacterium Xanthomonas Axonopodisis [20]. There are few illnesses as damaging as this one. Similarly, if a pepper plant is infected with blight, its yield might drop by 27.57.36 per cent. Fig. 2 depict a tomato leaf that has been infected with the blight disease.

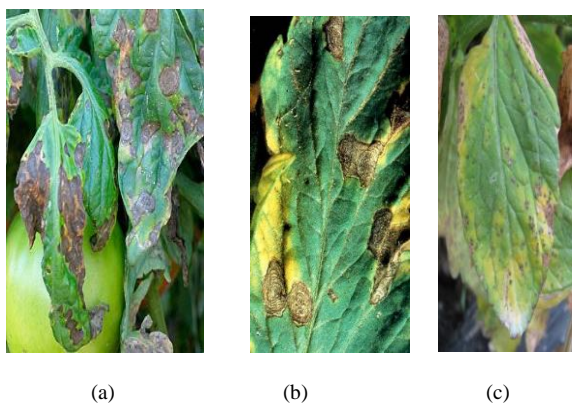


Fig. 2. Tomato leaf image.

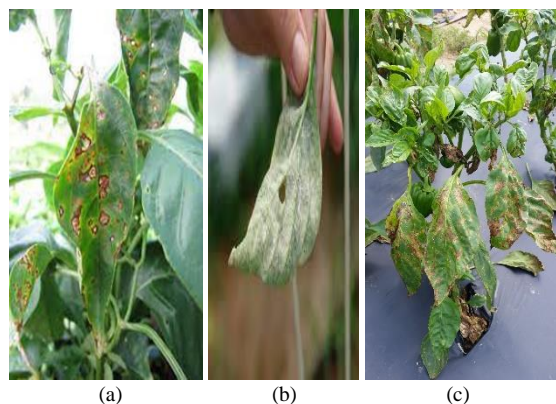


Fig. 3. Pepper leaf images.

Fig. 3 shows the pepper leaf images of the proposed work. Pepper leaf pictures, as well as unaffected input tomato, are taken into consideration for the input image, which is indicated by the letter I, and the impacted area of interest (ROI) is written as follows,

$$I = \sum_{x=1}^n S_x \quad (1)$$

In Eq. (1), n is represented as number of regions in input image, n number of regions are represented in terms of $S_1 + S_2 + \dots + S_n$.

Let us take two different regions then S_i not equal to S_j for $i \neq j$. The name for this kind of property is a disjoint property. A method for segmenting the intensity of histograms based on indices has been developed as a means of enhancing the segmentation and classification results produced by the aforementioned technique. Fig. 4 shows the proposed work block diagram.

1) *Denoising in input image:* This section explains the process of noise removal, since the provided input may have a possibility of having undesirable signals, which are referred to as noise. Denoising is a method that removes undesired signals from a picture while still preserving important information. Denoising is an essential step in the preprocessing of the picture, as it helps increase the accuracy of the final product. The median filter was used to eliminate

the sounds. The following is a representation of the three-by-three matrix that serves as the input to the median filter procedure. Fig. 5 shows the pixel selection and analysis of proposed work.

Let M be the input matrix, and then apply the sorting procedure to M in order to sort all of its values, and finally, get the median of M. In the previous illustration, the picture was smoothed down when the values of the surrounding neighborhoods were replaced with the median value of 214. After that, an adjustment was made to the image's contrast. The procedure described above is used in the process of pixel representation for the input picture in order to remove noise from the image and get the specific area of the leaf image that was damaged.

It can be seen in Fig. 6, that the noise has been reduced from the leaf photos. The median filter is applied to the photos of the three distinct illnesses, including foiled, rot, and rust, that have impacted the leaves.

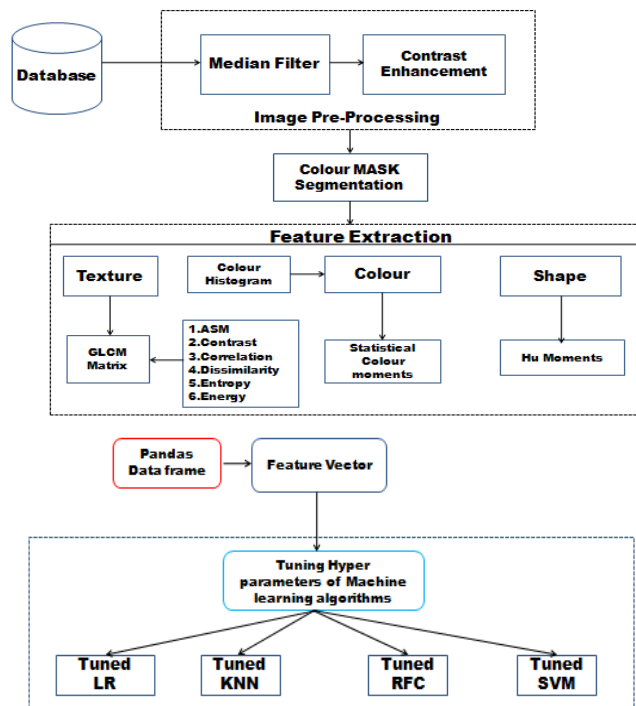


Fig. 4. Proposed first phase DLPD.

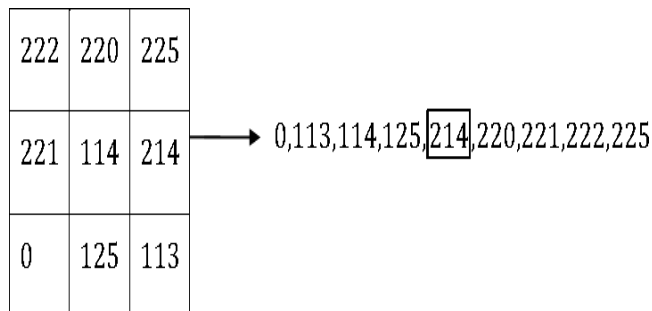


Fig. 5. Pixel selection.

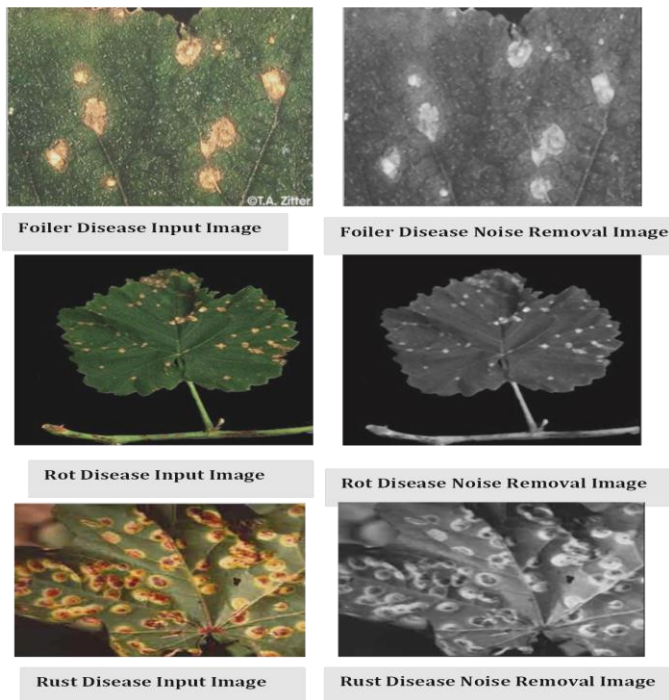


Fig. 6. Pre-processed leaf images.

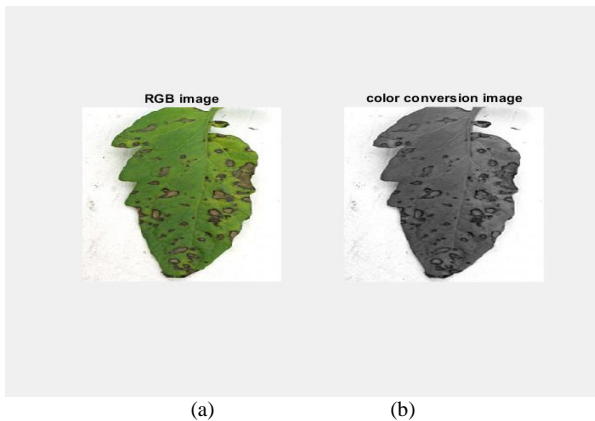


Fig. 7. (a). The original version, (b) Color conversion image.

Fig. 7 shows the original and color conversion results of provided dataset.

B. HSV Color Space Segmentation

The use of the Thresholding approach, the ROI may be extracted from the picture. Determine a value for the threshold, denoted by "T," depending on the highest and lowest points of the histogram. It is possible to employ the local threshold

method, the global threshold technique, or the optimum threshold technique depending on the application. The area that is of interest is known as the ROI, and it may either be a portion of the picture or the whole image. It depends on the application, but often it involves combining the OTSU threshold approach with the canny edge detection operator and dividing the jujube leaf into a number of regions. The function f is used to depict the picture of a Pepper leaf after it has been translated into a digital format using MATLAB (x, y) . By selecting the picture's threshold value, which is symbolized by the letter T, you can see that the image has been segmented. If the pixel intensity value is more than T, then it assigns value as 1; else, it takes the value as 0 for the pixel.

The formula for thresholding the leaf disease on Pepper is represented as follows, as in

$$\text{Threshold} = \begin{cases} 1 & \text{if } f(x, y) > T \\ 0 & \text{if } f(x, y) < T \end{cases} \quad (2)$$

Choose the T value but the result varies based on the selected domain. Same T value could not give accurate ROI for all kinds of input image. Implementation of the image segmentation technique used to separate the disease affected part from the background image. In this scenario first choose the seed point from the image. By using the seed point separate the diseased part. Choosing seed point is one of the biggest challenges in region growing method and wrong selection of seed point may lead to over segmentation. The study [21] implemented region growing method in cucumber downy mildew disease and get more accurate segmentation result compared with Otsu and K-means algorithm. Foliar disease of the leaf can be detected by using proposed region growing method and produced better result compared with existing algorithms. In this paper new algorithm is introduced based on the indices of histogram. During the analyze process, Pepper leaf images taken as input with the size of 256×256 as shown in Fig. 8.

Table II shows the feature metrics of the images and their values.

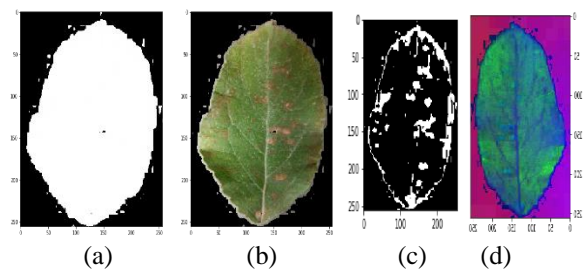


Fig. 8. (a) Graph cut algorithm (b) The original version, (c) Otsu thresholding, (d) HSV Color space segmentation.

TABLE II. COLOR MASKS FOR TOMATO, ONION, PEPPER LEAVES

S.no	Crop	Upper Green	Lower Green	Upper Brown	Lower Brown	Upper Yellow	Lower Yellow
1	Tomato	[86,255,255]	[36, 0, 0]	[30, 255, 200]	[8, 60, 20]	[40, 255, 255]	[21, 39, 64]
2	Onion	[86,255,255]	[36,0,0]	[30, 255, 200]	[8, 60, 20]	[255, 255, 255]	[3,3,3]
3	Pepper	[86,255,255]	[36,0,0]	[30, 255, 200]	[8, 60, 20]	[255, 255, 255]	[21,30,5]

C. Feature Extraction

1) *Color*: Extracting in all plant disease detection systems, color characteristics are essential since ill regions have different colors than healthy leaf photos; in this work, color features are recovered by measuring the color histogram and statistical color moments. The features of the HSV color histogram are extracted in the following stages: To begin, the cv2.COLOR_RGB2HSV function converts the input image from the RGB color space model to the HSV color space model. Quantizing the HSV color model has the effect of lowering both the cost of computation and the size of the feature. At this part of the process, each channel's H, S, and V components are quantized using eight different bins. The frequency distribution of quantized HSV values for each pixel in an image is shown by a histogram, which is created for each quantized picture. This histogram is specific to the image in question. In addition, statistical color moments are generated by first isolating the R, G, and B components of the input image, then computing the mean and standard deviation for each channel separately. This process is repeated for all three channels. In conclusion, color moments provide a feature vector that is six-dimensional.

2) *Shape*: An essential component of the quantification process for image objects is the extraction of shape information. In this work, form attributes are determined by calculating shape moments and locating shape edges. For the purpose of computing Hu moments, the function Hu moments found inside the Open – CV Python module is utilized. Converting the RGB image to grayscale is the first step that must be taken before attempting to calculate the Hu moments, which need just a single channel. Calculating the first 124 moments of the original picture is the responsibility of the CV2.moment module. After that, the real moments are input into the CV2 method, and the first six Hu moments are generated using the results of that. In order to get a shape feature vector, the Hu moments method and the array flattening method are used.

3) *Texture*: There are four distinct varieties of texture measurements that may be used in image processing, and they are structural, statistical, model-based, and transform-based. Statistical approaches may be used in this situation since the size of the texture is about equivalent to the size of the pixels. For instance, GLCM is an example of a statistical method that may be used to measure the textural features of an image, which can then result in various shades of grey. In order to generate the grey level co-occurrence matrix, fourteen Haralick texture characteristics, such as ASM, Correlation, Contrast, Entropy, Homogeneity, and Dissimilarity, in addition to eight additional features, were extracted. This process was carried out. This results in a thirteen-dimensional feature vector, with the fourteenth feature being omitted due to the substantial increase in the amount of processing it requires. Using manually chosen feature descriptors, the shape, color, and texture characteristics of the panda's data frame are

concatenated together to generate a 532-dimensional feature vector.

D. ML Classifier Tuning

The numerous layers that make up the CNN are capable of performing a wide variety of tasks, some of which include convolution, max pooling, activation, and a fully connected network (Fig. 9). Every CNN input image that is sent on to the Convolution layers includes filters, receptive fields, stride, padding, pooling, and ReLu stacks in some form or another. The receptive field of a CNN is the part of the network that monitors the activity of any filter that responds to each individual pixel. As further stack layers are added to the convolutional layers, it behaves in a linear manner throughout the process.

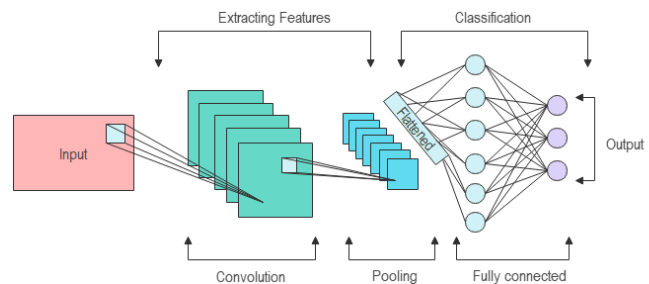


Fig. 9. Architecture of CNN.

When there is a negative number, the activation level that follows after it clips off to zero. This happens because negative numbers are less than zero. While the ReLu activation function converges more rapidly than the sigmoid activation function, it is also saturated in the negative region, which causes the gradient to be zero. This is because the ReLu activation function is a Convolutional Neural Network.

The very next pooling layer happens in between the convolutional layers, and as a consequence, it reduces the amount of data sampled by the network and gets rid of the other value data that is not relevant. After passing through the pooling layer, the dataset is shrunk down to a more manageable size, and the process of sending the dataset through the pooling layer is repeated until the desired output is produced. The matrix size will be decreased from 4x4 to 3x3 as a direct consequence of the maximum pooling layer, and then it will be decreased even more from there. The next layer is likewise fully connected, and it comprises neurons that have a full connection to all of the activations that were produced using a matrix multiplication. This layer is referred to as "totally linked."

Inside of a CNN network with several layers, the image is first warped and then converted into pixels. Following that, an activation of a neuron is performed for each location, and the results are compiled inside of the feature map. If the receptive field is moved one pixel further away from the activation layer, then the field plane will overlap with the previous activation by an amount that is equal to the field plane width minus one input value. This will occur if the receptive field is moved. After this, the fully connected layers will proceed to categories the many classes that have been instructed as binary values.

The supervised machine learning algorithms L.R., KNN, SVM, and R.F. are used in the initial step of the DLPDS process. In most cases, machine learning models are made up of two parameters: the default Model Parameter, and the hyperparameter.

Algorithm 1: Training and Testing phase

Step 1: Initialize the bias and weights to a random value.
Step 2: Input the training input vector and its targets.
Step 3: Compute the output of the hidden layer. The net input of the hidden layer is computed as follows,

$$y_k = \sum w_{ik}x_i + b_k$$

where b_k is the bias of the hidden layer, w_{ik} is the weights between the input and the hidden node and x_i is the input vector.

The output of the hidden layer is given by,

$$P_k = f(y_k)$$

where $f(0)$ is the activation function of the neuron.

Step 4: Compute the output of the output layer. The net input of the output layer is computed as follows.

$$z_1 = \sum w_{jk}p_k + b_i$$

where b_1 is the bias of the output layer, w_{jk} is the weights between the hidden and the output node and p_k is the output vector from the hidden neuron. The output of the output layer is given by.

$$y_1 = f(z_1)$$

where $f(0)$ is the activation function of the neuron.
Step 5: Compute the error at the output layer and at the hidden layer.

The error at the output layer is given by,

$$E_j = (t_j - y_j)f'(z_j)$$

The error at the hidden layer is given by.

$$E_k = \sum \delta_j w_{jk} f'(y_k)$$

Step 6: Update and renew the weights and bias with the learning rate α for output layer which is given by,

$$\Delta w_{jk} = \alpha E_j x_i$$

$$\Delta b_k = -\alpha E_k$$

$$w_k(\text{new}) = w_{ik} + \Delta w_{ik}$$

$$b_k(\text{new}) = b_j + \Delta b_k$$

Update and renew the weights and bias with the learning rate α for hidden which is given by,

$$\Delta w_{ik} = \alpha E_k y_k$$

$$\Delta b_{kk} = -\alpha E_j$$

$$\Delta w_{jk}(\text{new}) = w_{jk} + \Delta w_{jk}$$

$$b_j(\text{new}) = b_j + \Delta b_j$$

Step 7: Repeat step 2 to step 6 until the stopping criteria is reached.

5.2.2 Testing Process

Step 1: Feed the unknown data vector X
Step 2: Compute the output of the hidden layer. The net input of the hidden layer is computed as follows,

$$y_k = \sum w_{ik}x_i + b_k$$

The output of the hidden layer is given by,

$$P_k = f(y_k)$$

Step 3: Compute the output of the output layer. The net input of the output layer is computed as follows.

$$Z_1 = \sum w_{jk}P_k + b_i$$

The output of the output layer is given by.

$$y_j = f(Z_j)$$

where $f(0)$ is the activation function of the neuron.

Step 4: Obtain the classification result from the output neuron in the output layer.

The practitioner may modify the hyperparameters to improve the classification results. Grid search and Manual search are used in this study for the Tomato, Tomato, and Pepper datasets because they provide results more quickly than other techniques. Fig. 10 shows the proposed methodology block diagram stage 2.

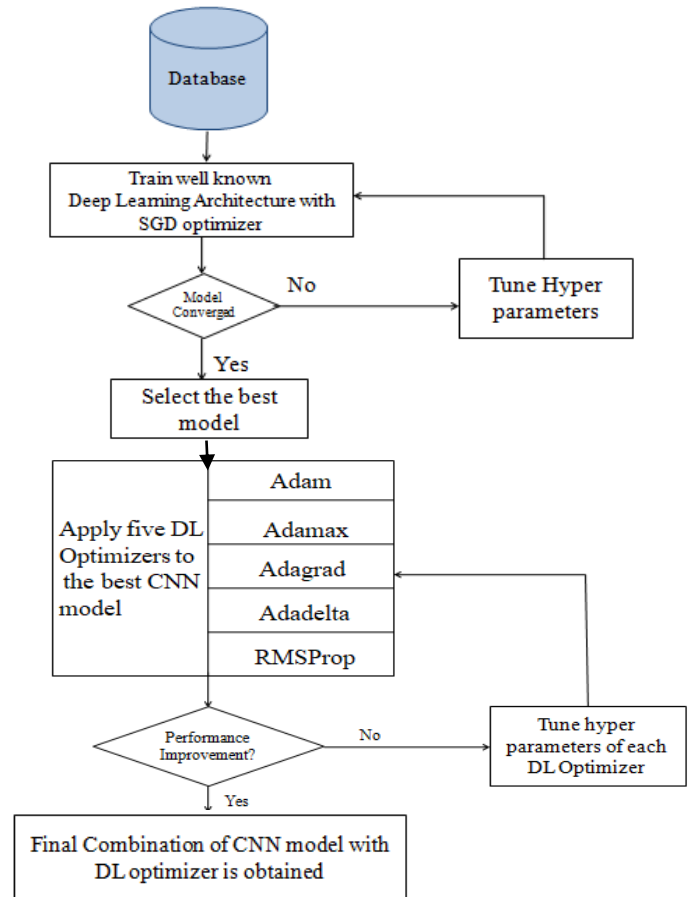


Fig. 10. Proposed second phase DLPDS.

E. Deep Learning Optimizers

First, the SGD optimizer is put to use in order to train the previously stated well-known architectures on the basis of validation accuracy, after which the best CNN model is selected. In conclusion, the Six DL Optimizers, whose features are outlined in Table III, are fine-tuned in order to improve the classification accuracy.

TABLE III. PARAMETERS OF DEEP LEARNING

gamma	alpha	value	Accuracy %		
			Tomato	Pepper	Onion
Scale	0.1-0.3	alpha	0.803333	0.9862	0.866667
		alpha	0.783667	0.9921	0.838333
		alpha	0.711667	0.9752	0.838333
	0.4-0.6	alpha	0.832667	0.9923	0.821667
		alpha	0.809667	0.9775	0.729
		alpha	0.741333	0.9921	0.884667
	0.7-1.0	alpha	0.881333	0.9712	0.871333
		alpha	0.866667	0.9923	0.729
		alpha	0.803333	0.9862	0.866667

IV. EXPERIMENTAL ANALYSIS

For the purpose of this study, ten thousand photographs spanning fourteen distinct categories were taken in the area of agriculture and sourced from the EPPO Global Database, both of which are open to the public. In addition, there are eight other categories that make up the Classification of Tomato plant diseases. In this study, we utilize datasets including both pepper and tomato. In both the pepper and tomato, onion datasets, the pictures are divided into training and testing sets in an 80:20 arrangements. There are a total of 5932 photos in the pepper and tomato collection shown in Table IV.

A. Tuning ML Classifiers

1) *LR Tuning*: C value [100, 10, 1.0, and 0.1], solver ['newton-cg,' 'lbfgs,' and 'liblinear,'] and penalty [12] are the most important parameters that are extracted from the default Python specification to tailor the performance of LR. The settings shown above have been applied to these parameters in order to modify the Tomato, Pepper, and Tomato characteristics shown in Table IV.

TABLE IV. LR-TUNING RESULTS

Sigma Value	Penalty	Solver	Accuracy %		
			Tomato	Pepper	Onion
100	elasticnet	saga	0.936333	0.937667	0.938667
		sag	0.936333	0.939000	0.937333
		liblinear	0.945333	0.929000	0.974667
10		lbfgs	0.874561	0.894561	0.925333
		newton-cg	0.861234	0.929000	0.925333
		lbfgs	0.797531	0.974333	0.895333
1.0		liblinear	0.971667	0.945333	0.970667
		sag	0.974000	0.974333	0.974000

2) *SVM Tuning*: The kernels- ['linear', 'poly', and 'rbf'] parameter, the C value- [10, 1.0, and 0.1] parameter, and the gamma-scale parameter are the ones that were chosen from the default Python SVM classifier specification. Table V demonstrates that an SVM with a C value of 1.0 and a kernel of poly achieves an accuracy of 88.13% for tomato features, 97.12% for Pepper features, and 87.13% for onion leaf features.

Table V shows the existing tuning values of the SVM classifier.

TABLE V. SVM-TUNING RESULTS

ML classifier	Para meters	Tomato								Accuracy %
		Healthy	Early blight	Leaf mold	Septoria leaf spot	Target spot	Bacterial blight	Late blight	Spider millets	
TLR	P	0.96	0.79	0.84	0.80	0.81	0.94	0.81	0.89	85.50
	R	0.95	0.81	0.91	0.75	0.88	0.91	0.75	0.88	
	F1	0.96	0.80	0.88	0.77	0.84	0.93	0.78	0.88	
TSVM	P	0.92	0.82	0.88	0.81	0.79	0.93	0.82	0.87	85.62
	R	0.92	0.81	0.94	0.76	0.86	0.92	0.81	0.83	
	F1	0.92	0.81	0.91	0.79	0.82	0.93	0.82	0.85	
TKNN	P	0.92	0.80	0.82	0.84	0.72	0.93	0.90	0.73	82.87
	R	0.86	0.74	0.95	0.74	0.79	0.84	0.63	0.82	
	F1	0.85	0.75	0.85	0.76	0.74	0.88	0.72	0.80	
MCNN	P	0.97	0.85	0.94	0.92	0.77	0.97	0.87	0.92	90.12
	R	0.98	0.88	0.96	0.88	0.90	0.93	0.80	0.88	
	F1	0.98	0.86	0.95	0.90	0.82	0.94	0.84	0.90	

3) *KNN Tuning*: The parameters n neighbors [3, 4], metric ['Euclidean', 'Manhattan'], and weights ['uniform', 'distance'] were chosen from the KNN default python specification. According to Table VI, using n neighbors =4 with metric = Manhattan, weights = Distance results in an accuracy of 82.75% for tomato, 90.12% for tomato, while using n neighbors =3 with Euclidean metric and distance as weight results in an accuracy of 99.21% for Pepper leaf characteristics.

4) *RF Tuning*: When it comes to tuning, the parameters max features ['sqrt', 'log2'], and n estimators [10, 100, and 1000] were chosen from the RF default Python specification. According to Table VII, using max features = sqrt and n estimators =1000 results in an accuracy of 95% for tomato and 99.50% for Pepper. Yet, using the same max features and the same number of estimators results in an accuracy of 90.12%.

TABLE VI. KNN-TUNING RESULTS

Max-Features	metric	weights	Accuracy %		
			Tomato	Pepper	Onion
10	sqrt	estimators	0.7975	0.9	0.9862
		estimators	0.81	0.891	0.9921
	log2	estimators	0.8062	0.9051	0.9752
		estimators	0.82	0.8937	0.9923
100	sqrt	estimators	0.7762	0.8811	0.9775
		estimators	0.8087	0.8957	0.9921
	log2	estimators	0.8025	0.8912	0.9712
		estimators	0.8275	0.9012	0.9923

B. First Phase DLPDS with Tuned Models

1) *Tomato diseases*: For the purpose of categorization, a total of 4000 different photos of tomato leaves are taken, 500 of which are healthy and 3500 of which display one of seven different disorders. For the objectives of experimentation, the 50:50 technique has been selected. The training and testing photos consist of 250 healthy and 1750 sick images.

TABLE VII. RF-TUNING RESULTS

max-features	estimators	Accuracy %		
		Tomato	Pepper	Onion
sqrt	10	82.50	91.00	99.12
	100	90.12	93.00	99.21
	1000	89.50	95.00	99.50
log2	10	82.00	91.00	98.96
	100	88.00	92.23	99.34
	1000	88.62	93.50	99.37

SHAPE and TEXTURE are put to use for the purpose of distinguishing sick samples and healthy samples using 10-fold cross-validation (a). According to Table VIII, the Tuned Random Forest Classifier has an overall accuracy of 90.12%. Higher values of performance measures include Precision

equal to 0.97 for both healthy tomatoes and tomatoes with bacterial blight, recall equal to 0.98 and 0.96, and F1 score equal to 0.98 and 0.95 for both healthy tomatoes and healthy tomatoes with leaf mold shown in Table VIII.

TABLE VIII. PERFORMANCE METRICS BASED ON ACCURACY

Sigma Value	Penalty	Solver	Accuracy %		
			Tomato	Pepper	Onion
100	elasticnet	saga	0.936333	0.937667	0.938667
		sag	0.936333	0.939000	0.937333
		liblinear	0.945333	0.929000	0.974667
10		lbfgs	0.874561	0.894561	0.925333
		newton-cg	0.861234	0.929000	0.925333
		lbfgs	0.797531	0.974333	0.895333
1.0		liblinear	0.971667	0.945333	0.970667
		sag	0.974000	0.974333	0.974000

Moreover, there are two classifications that are used for categorizing tomato leaf diseases, and each class has a selection of one thousand photos. The majority of the illnesses that might affect tomato seem to be bacterial infections, according to the visual examination. Hence, any and all photographs of a sick tomato plant are saved under the heading "tomato diseases," whereas any and all images of a healthy tomato plant are saved under the heading "healthy tomato" shown in Table IX.

The tomato train and the test dataset are both developed by the use of the 50:50 approach. Validation of the tomato test picture dataset was accomplished with the help of the tuned four ML models. MCNN delivers 95% overall accuracy with Precision= 0.96, F1 score= 0.95 for bacterial illness characteristics, and Recall =0.96 for Healthy leaf features. The experimental findings are produced by ten-fold cross-validation, as shown in Fig. 11(b), and from Table IX, MCNN gives 95% overall accuracy.

TABLE IX. PERFORMANCE METRICS OF TOMATO LEAVES

ML classifier	Parameters	Tomato		
		Bacterial Disease	Healthy	Accuracy %
TLR	P	0.90	0.87	88.50
	R	0.86	0.91	
	F1	0.88	0.89	
TSVM	P	0.95	0.91	92.50
	R	0.90	0.95	
	F1	0.92	0.93	
TKNN	P	0.94	0.87	90.50
	R	0.86	0.95	
	F1	0.90	0.91	
TRF	P	0.96	0.94	95
	R	0.94	0.96	
	F1	0.95	0.95	

2) *Pepper diseases*: There are four categories that make up the Pepper leaf disease categorization system, and each category has a thousand pictures. A total of four thousand photographs are amassed, and then, applying the 50:50 technique, the picture datasets for the train and test are partitioned into healthy images, class 500 images, and test images, respectively. There are 1500 photos that fall under the heading of the sick image, which may be used for training and testing purposes. The test picture dataset was verified by using the extracted characteristics of Pepper leaves as well as the results of ten-fold cross-validation, which are shown in Fig. 11(c). Table X shows that the majority of the Tuned models got improved classification results, with TKNN (Tuned K Nearest Neighbor) reaching an accuracy of 99.25% and MCNN achieving a greater accuracy of 99.50.

TABLE X. PERFORMANCE METRICS OF PEPPER LEAVES

ML classifier	Parameter	Pepper				Accuracy %
		Healthy	Cercospora_1 eaf_spot	Common rust	Northern_Leaf_Blight	
TLR	P	1.00	0.95	1.00	0.90	96.25
	R	1.00	0.90	1.00	0.96	
	F1	1.00	0.93	1.00	0.93	
TSVM	P	1.00	0.97	1.00	0.92	97.12
	R	1.00	0.92	1.00	0.97	
	F1	1.00	0.94	1.00	0.94	
KNN	P	1.00	0.97	1.00	1.00	99.25
	R	1.00	1.00	1.00	0.97	
	F1	1.00	0.99	1.00	0.98	
TRF	P	1.00	0.99	1.00	0.99	99.50
	R	1.00	0.99	1.00	0.99	
	F1	1.00	0.99	1.00	0.99	

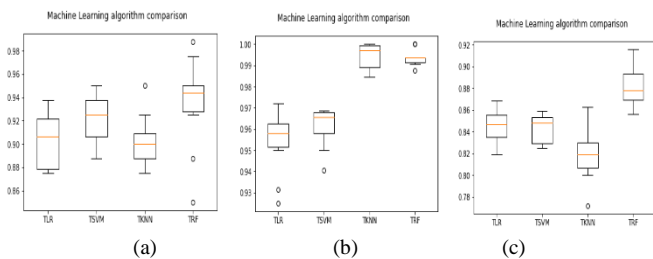


Fig. 11. Ten-fold cross-validation results of (a) Onion, (b) Tomato, (c) Pepper.

C. Second Phase DLPDS

The Tomato, Tomato, and Pepper leaf pictures were used to train the chosen DL Architectures and the experimental findings were represented by validation, training accuracy/loss, Precision, F1 Score, and Recall. The model or optimizer that achieves the highest possible Validation accuracy and F1 score is taken into consideration to be the best option for the Second Phase DLPDS that is being suggested. The accuracy and loss measurements in the training and validation stages need a combined total of 20 epochs in order to converge.

1) *Performance of pretrained models*: Accuracy, sensitivity, selectivity, kappa coefficient, and mean square

error are some of the several performance metrics that are taken into consideration. The following is a representation of the notations that are used in the computation of the metrics: TP - true positive, TN - true negative, FN - false negative, TP - positive, TN - negative, FN - false negative. The following is the calculation for the performance metrics:

- Accuracy: It is the measure of how well the classifier correctly identifies whether the leaf is healthy or diseased.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}} \quad (3)$$

- Sensitivity: It is also known as recall and it represent the measure of proposition of diseased leaf correctly identified as such.

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

- Specificity: It is also termed by True Negative Rate, and it represents the proportion of healthy leaf correctly identified.

$$\text{Specificity} = \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}} \quad (5)$$

- Kappa Coefficient: The kappa coefficient is used to determine how closely the original values and the graded values are related to one another. A kappa score of one shows that all respondents are in complete agreement, while a value of 0 indicates that there is no consensus. Here is how it is figured out,

$$\text{Kappa} = \frac{z \sum_{j=1}^n m_{jj} - \sum_{j=1}^n (G_j C_j)}{z^2 - \sum_{j=1}^n (G_j C_j)} \quad (6)$$

where j is the number of the class, Z is the total number of graded values that are compared to the original values, m (i,j) is the number of values belonging to the truth class j that have been classified as class j, Cj is the total number of expected values belonging to class j, and Gj is the total number of truth values belonging to class j. where j is the number of the class, Z is the total number of graded values that are compared

During the second phase, RT is used to pinpoint the precise location of the leaf illness as well as determine its degree of severity shown in Table XI.

TABLE XI. COMPARISON OF PROPOSED TRAINING FUNCTION AND TRAINBR

Sl.No	Training Function	MSE	Accuracy	Kappa Coefficient
1	Trainbr	0.03	93.4	0.892
2	Proposed Method	0.02	95.64	0.928

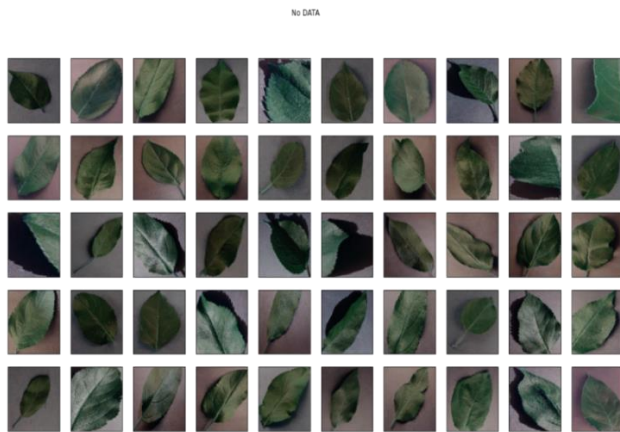
The categorized picture is then put through RT in order to pinpoint the area that is affected by the illness. If the picture is determined to be a disease, a morphological operation is performed on it, coupled with the determination of an appropriate threshold, in order to differentiate the sick area from the backdrop. The segmented output is created by applying a morphological opening operation with a square structural element to the identified picture. This results in the

output being segmented. The RT is applied to the sick leaf that has been segmented. The results of the morphological operation are first segmented, and then the RT is applied to the results of that segmentation. The RT output provides evidence that the disease node is present in the leaf. The radon transform is useful for pinpointing exactly where the sick area is located in the body.

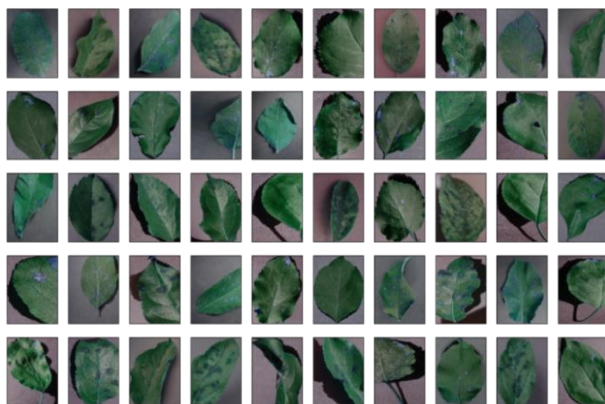
There is no indication of either underfitting or overfitting in the experimental findings of the pre-trained models using the Pepper picture dataset. From Fig. 12(b) Enhanced GoogleNet was able to achieve the highest possible level of Validation accuracy by making use of the idea behind the Inception module. The Improved GoogleNet and SGD optimizer, together with a comparative study of the several models shown in Table XII, led to the best possible F1 score of 0.993.

TABLE XII. PERFORMANCE OF PRE-TRAINED MODELS

Parameters	MobileNetV2	Improved GoogleNet
Training Accuracy	98.99	98.99
Validation Accuracy	97.73	99.92
Training Loss	0.220	0.021
Validation Loss	0.285	0.002
Precision	0.982	0.991
F1 Score	0.986	0.993



(a): No DATA.



(b) Yes DATA.

Fig. 12. Training and testing data.

TABLE XIII. ANALYSIS OF SEGMENTATION

Raw Input Image	K-Means Clustering	Fuzzy Logic	Region-Based Segmentation	HSV Color Segmentation
Tomato blight disease	0.684	0.82	0.8843	0.8368
Tomato Leaf Spot	0.66	0.856	0.801	0.8465
Tomato Powdery Mildew	0.68	0.861	0.8154	0.8624
Pepper blight disease	0.803	0.865	0.8266	0.868

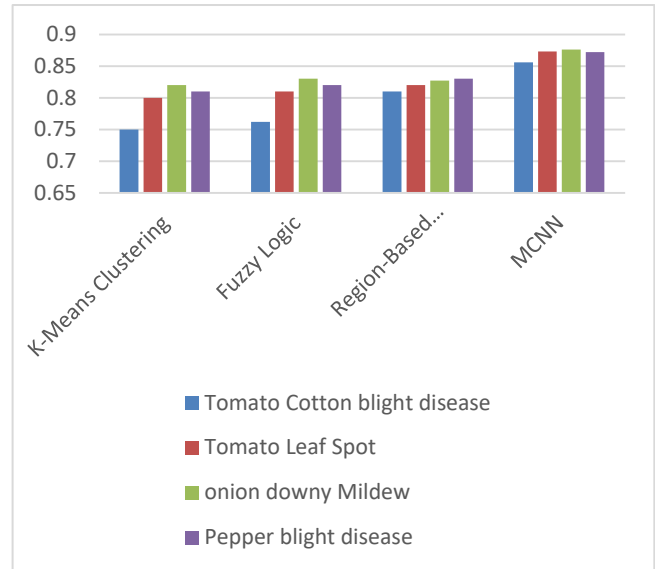


Fig. 13. DSC Performance of segmentation techniques.

Fig. 13 makes it abundantly evident that the indices-based intensity histogram segmentation technique assures a somewhat superior performance value for various plant leaves, such as those affected by tomato blight disease, tomato spot, tomato powdery mildew, and Pepper blight disease, respectively. The divided components each provide important information that may be utilized to examine the characteristics and facts connected to the condition. Table XIII presents the results of the mutual information calculation for the segmented area.

TABLE XIV. MUTUAL INFORMATION FOR SEGMENTATION RESULT

Raw Input Image	K-Means Clustering	Fuzzy Logic	Region-Based Segmentation	MCNN
Tomato blight disease	0.67	0.75	0.83	0.85
Tomato Leaf Spot	0.68	0.78	0.86	0.86
Tomato Powdery Mildew	0.73	0.80	0.88	0.868
Pepper blight disease	0.76	0.83	0.82	0.878

According to what is shown in Table XIV. It is abundantly obvious that the indices that are based on the intensity of the histogram include various information about impacted diseases in an efficient way. Indices Based Intensity Histogram Segmentation approach segment region contain significant information because it assures the largest mutual information value. This can be understood since the technique for indices-

based intensity histogram segmentation segments the histogram based on its intensity. Fig. XIV is a graphical depiction of the linked data for the mutual information value, and it was created using Table XIV.

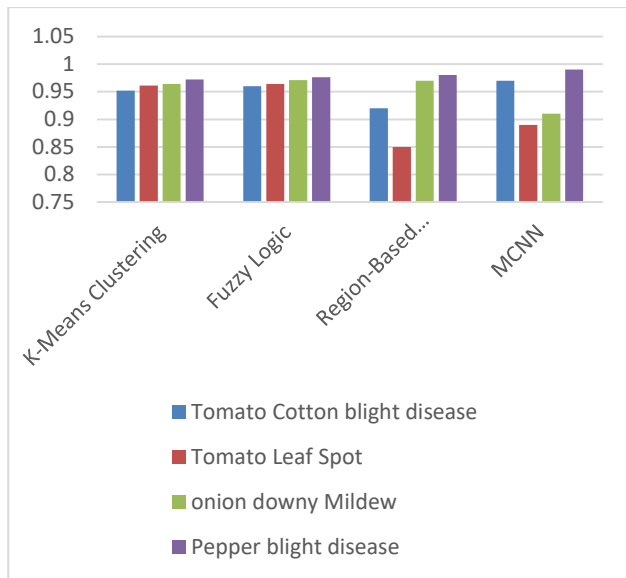


Fig. 14. Mutual information for different segmentation methods.

According to Fig. XIV, a technique to segmenting intensity histograms that is based on indices assures a high mutual information value for various plant leaves, such as those affected by tomato blight disease, tomato spot, tomato powdery mildew, and Pepper blight disease. The divided components each provide important information that may be utilized to examine the characteristics and facts connected to the condition. Effective retrieval of disease-related information from the segmented area is accomplished, and the accuracy of the segmented region's diagnostic performance is evaluated by means of the sensitivity and specificity metrics that are shown in Table XIV.

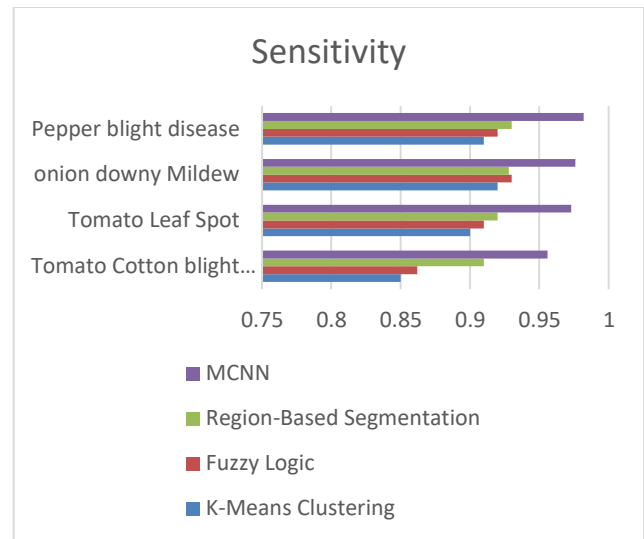
Fig. 15(a) and (b) shows that using a strategy that segments intensity histograms based on indices assures a high sensitivity and specificity value for various plant leaves, such as those affected by tomato blight disease, tomato spot, tomato powdery mildew, and Pepper blight disease. The segmented sections include valuable information that can be used to assess the disease-related characteristics and information that can be used to obtain disease-related information with the maximum possible accuracy, as seen in Table XV.

The results shown in the preceding Table XV make it abundantly evident that the indices based on the intensity histogram segmented area include a high degree of accuracy (88.78%) about the afflicted illness portion in an efficient way.

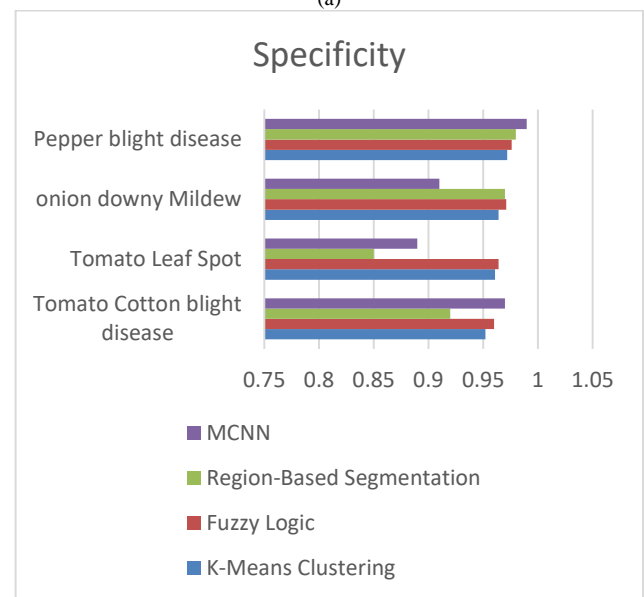
2) Discussion: The experimental results show that the enhanced versions of GoogleNet and MobileNetV2 using SGD Optimizer with ImageNet weights achieve a significantly more acceptable degree of validation accuracy. These two models are compared using five different optimizers and the same amount of epochs in an effort to improve classification

accuracy. Based on the data in Table XI, we may make the following inferences:

Training the pre-trained models with various optimizers led to notable improvements in validation, training accuracy/loss, F1 Score, precision, and recall.



(a)



(b)

Fig. 15. (a) Sensitivity and (b) Specificity.

TABLE XV. ACCURACY FOR SEGMENTATION RESULT

Raw Input Image	K-Means Clustering	Fuzzy Logic	Region-Based Segmentation	MCNN
Tomato blight disease	82.1	83.5	86.34	87.31
Tomato Leaf Spot	82.8	83.88	85.134	87.543
Tomato Powdery Mildew	83.1	84.21	86.43	88.46
Pepper blight disease	83.78	84.72	86.83	88.78

The best validation accuracy may be achieved with the use of optimizers such as Adam, SGD, and Adadelta.

- Enhanced GoogleNet with Adam Optimizer achieved a 99.95% success rate and an F1 score of 0.997, demonstrating the efficacy of the proposed tuning strategy. The results of this experiment are also significantly improved upon when compared to those of previous studies. Therefore, the given technique may be used to a wide range of plant diseases.
- The F1 score and validation accuracy for MobileNet were both improved by using the Adam and Adadelta optimizers.

However, it is possible to see a decrease in performance after switching from SGD to RMSProp and Adamax for improved GoogleNet and MobileNetV2. This is the case despite the fact that these algorithms are designed to improve performance.

V. CONCLUSION AND FUTURE WORK

Plant disease and pest attack is a major threat to the farmers and the farming industries. It impacts the economy majorly by destroying the plants and production quality. An automatic plant disease detection system is a key factor in the growth and benefit of farm production. This research is carried out based on the image-based automatic disease detection system which includes various image processing and CNN techniques. The proposed methods are developed to enhance the detection system and classify diseases. Training, validation, and testing were conducted on various publicly available datasets. The datasets consist of images captured under different lighting conditions, resolution, position, and complex background to train the system with all possible complexities to avoid the misclassification rate. When compared to ML models, the F1 Score and Accuracy of the Enhanced GoogleNet and MobileNetV2 models were shown to be superior. The best results for differentiating between photos of diseased and healthy tomato leaves were achieved using the DL model and optimizer combination known as Improved GoogleNet with the Adam optimizer, which achieved a validation accuracy of 99.5% and an F1 score of 0.997%. Furthermore, two distinct DL architectures were tweaked with five distinct DL Optimizers to enhance classification accuracy. The future research in the automatic plant disease and pest detection method focuses mainly on increasing the efficiency of the system by reducing the computational time. It also further focuses on treatment and prevention methods based on the severity of the impact. The entire system should be implemented in real-time using mobile and web applications, where the test images can be taken using any device including mobile cameras, drone cameras, satellite images and sensor images. The real-time applications store and retrieve all the information based on cloud services. Which enables the farmers to operate the farming portal to detect the disease or pest attack based on the captured images and based on the impact and severity of the disease the portal provides a solution to the farmer by prescribing the procedure for the treatment and prevention techniques like cross farming, seedling selection, ground detection, canopy estimation, water level detection,

phenotype, and genotype evaluation to avoid the further occurrence of the same disease or pest attack.

ACKNOWLEDGMENT

The author would like to thank Deanship of Scientific Research at Majmaah University for supporting this work under Project Number No. R-2023-536.

REFERENCES

- [1] A.S.Deshapande, S.G.Girardi, K.G.Karibasappa and S.D.Desai, "Fungal disease detection in Pepper leaves using Haar wavelet features," *Information and Communication Technology for Intelligent Systems*, vol.106, pp.275-286, 2019.
- [2] X.Zhang, Y.Qiao, F.Meng, C.Fan and M.Zhang, "Identification of Pepper leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol.6, pp.30370–30377,2018.
- [3] K.Golhani, S.K.Balasundram,G.Vadamalai and B.Pradhan, "A review of neural networks in plant disease detection using hyperspectral data airborne imaging spectrometer for applications," *Information processing in Agriculture*, vol. 5,no.3,pp. 354-371, 2018.
- [4] J.G.ArnalBarbedo, "Plant disease identification from individual lesions and spots using deep learning," *Biosystems Engineering*, vol.180, pp.96 - 107,2019.
- [5] Z. Xihai, S. Yue Qiao, V. FanfengMeng, C. Chengguo Fan and A. Mingming Zhan, "Identification of Onion leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol 6, pp. 30370 – 30377, 2018.
- [6] P.Sethy, S.Gouda, N.Barpanda and A.Rath, "Detection of white ear-head of rice crop using image processing and machine learning techniques," *Smart Computing Paradigms: New Progresses and Challenges*, vol.1,pp. 87-95, 2020.
- [7] P. Patrick Wspanialy and C. MedhatMoussa, "A detection and severity estimation system for generic diseases of tomato greenhouse plants," *Computers and Electronics in Agriculture*, vol.178, 105701, 2020.
- [8] X.E.Pantazi, D.Moshou and A.A.Tamouridou, "Automated leaf disease detection in different crop species through image features analysis and One-Class Classifiers," *Computers and Electronics in Agriculture*, vol.156, pp.96–104,2019.
- [9] J.Chen, H.Shao and C.Hu, "Image segmentation based on mathematical morphological operator," *Colorimetry and Image Processing*, doi:10.5772/intechopen.72603,2018.
- [10] P.Xu, G.Wu, Y.Guo, X.Chen, H.Yang et al., "Automatic wheat leaf rust detection and grading diagnosis via embedded image processing system," *Procedia Computer Science*, vol.107, pp.836–841,2017.
- [11] A.Cruz, Y.Ampatzidis, R.Pierro and A.Panattoni, "Detection of grapevine yellows symptoms in vitis vinifera with artificial intelligence," *Computers and Electronics in Agriculture*, vol.157, pp.63–76,2018.
- [12] E. Proposed workjie Liu, P. Yongjun Zhang, B. Haisheng Fan, Z. Yongjie Zou and Q. YongbinQin, "Detection of late blight in potato leaves based on multifeature and svm classifier," *Journal of Physics: Conference Series*, vol.1518, no.012045,2020.
- [13] E.Alehegn, "Ethiopian Onion diseases recognition and classification using support vector machine," *International Journal of Computational Vision and Robotics*, vol. 9,no.1, pp.90 – 109. 2019.
- [14] M.Sharif, M.Khan and M.Javed, "Detection and classification of citrus diseases in agriculture based on optimized proposed workighted segmentation and feature selection," *Computers and Electronics in Agriculture*, vol.150, pp.220–234,2018.
- [15] S. Solemane Coulibalya, V. Bernard Kamsu-Foguema, C. DantoumaKamissokob and T.DaoudaTraore. "Deep neural networks with transfer learning in millet crop images," *Computers in Industry*, vol.108, pp. 115–120,2019.
- [16] K.Bashir, M.Rehman and M.Bari "Detection and classification of rice disease an automated approach using textural features," *Mehran University Research Journal of Engineering & Technology*, vol. 38, no. 1, pp.239-250,2019.

- [17] J.Pujari, R.Yakkundimath and A.Byadgi, "Classification of fungal disease symptoms affected on cereals using color texture features," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol.6, no.6, pp.321-330, 2013.
- [18] A.Das, R.Dutta, S.Das and S.Sengupta, "Feature selection using graph-based clustering for rice disease prediction," Computational Intelligence in Pattern Recognition, pp. 589–598, 2020.
- [19] Baljon, M.; Sharma, S.K. Rainfall Prediction Rate in Saudi Arabia Using Improved Machine Learning Techniques. Water 2023, 15, 826.
- [20] A.Chatterjee, S.Roy and S.Das, "Feature selection using rough set theory from infected rice plant images," Computational, pp. 417–427, 2020
- [21] R.Yadav, Y.Rana and S.Nagpal, "Plant leaf disease detection and classification using particle swarm optimization," Machine, pp 294-306, 2019.

Lung Cancer Classification using Reinforcement Learning-based Ensemble Learning

Shengping Luo*

College of Physics and Electronic Information, Nanchang Normal University
Nanchang 330000, Jiangxi, China

Abstract—Lung cancer is a significant health issue affecting millions of people worldwide annually. However, current manual detection methods used by physicians and radiologists to identify lung nodules are inefficient because of the diverse shapes and locations of the nodules in the lungs. New methods are needed to improve the accuracy and speed of detecting lung nodules. This is important because early detection of nodules can increase the likelihood of successful treatment and recovery. This paper introduces a new LLC-QE model that combines ensemble learning and reinforcement learning to classify lung cancer. Initially, the model undergoes pre-training through the utilization of the Artificial Bee Colony (ABC) algorithm. This approach aims to decrease the probability of the model getting stuck in a local optimum. Subsequently, a set of convolutional neural networks (CNNs) is used to simultaneously derive feature vectors from input images, which are subsequently combined for classification in downstream processes. The LIDC-IDRI dataset, predominantly composed of cases without cancer, was employed to train and evaluate the model. To mitigate the dataset imbalance, the training procedure using reinforcement learning is formulated as a series of interconnected decisions. During this process, the images are regarded as states; the network acts as the agent, and the agent is given a greater reward/punishment for accurately/incorrectly classifying the underrepresented class compared to the overrepresented class. The LLC-QE model achieves excellent results (F measure 89.8%; geometric mean 92.7%), outperforming other deep models. Identifying the optimal values for the reward function and determining the ideal number of CNN feature extractors in the ensemble are achieved through experiments conducted on the study dataset. Ablation studies that exclude ABC pre-training and reinforcement learning from the model confirm these components' independent positive incremental impact on the model's performance.

Keywords—Lung cancer; ensemble learning; reinforcement learning; artificial bee colony; convolutional neural network

I. INTRODUCTION

In recent years, the global mortality rate for lung cancer has risen significantly. This indicates that lung cancer has emerged as among the deadliest forms of cancer in recent decades [1]. However, over 50% of lung cancers can be treated successfully if detected early [2, 3]. Automatic cancer detection can significantly reduce the time required for diagnosis, leading to timely treatment. Sufficient forms of lung cancer are not visible to the naked eye, making automated diagnosis a valuable tool in reducing human error [4]. The computer-aided diagnosis (CAD) system can assist radiologists in rapidly and precisely detecting and diagnosing abnormalities. This can aid in

identifying and diagnosing lung cancer at an earlier stage, resulting in more effective treatment options [5].

Computed tomography (CT) is a widely used method for detecting lung cancer, leading to an increase in CT images and putting pressure on radiologists [6]. To ease this burden, Computer-Aided Diagnosis (CAD) systems have been developed to aid in nodule detection [7, 8]. Detecting nodules is a complex task given their various sizes, shapes, and positions. Deep learning, particularly in CAD, has shown potential to enhance nodule detection. Examples include ZNET [9] using the U-Net architecture [10], Resnet utilizing a 3D CNN, and JianPeiCAD [11], which employs a multi-scale rule-based approach followed by a broad-channelled 3D CNN. While 3D CNNs capture CT scans' details, they come with longer training times and higher storage needs. The varying slice thickness in CT scans complicates 3D imaging, making 2D imaging more suitable in terms of training duration and resource use, making it a preferred method for nodule identification.

Imbalanced class distribution is a pressing issue in deep learning, especially in lung cancer classification, where the uneven spread between positive and negative cases hampers model accuracy [12]. To address this, data-level methods like over-sampling and under-sampling are employed. Over-sampling, such as Synthetic Minority Oversampling Technique (SMOTE) [13], creates synthetic examples for the minority class, while under-sampling techniques like NearMiss [14] reduce majority class instances. However, these can lead to overfitting or loss of vital data. Algorithm-level solutions amplify the minority class's influence using ensemble learning, cost-sensitive methods, and decision threshold adjustments. Cost-sensitive techniques assign different misclassification costs, ensemble methods utilize multiple classifiers, and threshold adjustments modify the classification threshold during tests. Some deep learning strategies focus on learning distinct features in unbalanced data or ensure balanced mini-batch training in convolutional networks. These approaches aim to improve classification precision in the face of imbalanced data [15].

In the past several years, deep reinforcement learning (DRL) [16] demonstrated successful applications in various areas, including computer games, robot control, and recommendation systems. DRL helps in removing noisy data and enhancing features, which ultimately boosts the performance of the classification system [17]. The classification process is a sequential decision-making task that requires the acquisition of an optimized policy. However, the

computational time required for the process is amplified due to the elaborate simulations conducted between agents and environments. Some researchers have utilized deep reinforcement learning to learn valuable data features and enhance the useful features of the classifier [18-21]. An ensemble pruning approach has also been developed, which selects the best sub-classifiers with the help of RL, which is effective for small data [22]. However, there has been a minimal investigation into using DPL in imbalanced classification, particularly in medical images. DPL is suitable for imbalanced classification as it rewards or penalizes the minority class more to attract more attention.

ABC [23] is utilized for optimization, inspired by the manner in which honeybees hunt for food. The algorithm imitates the way bees search for food sources by using three components: employed bees, onlooker bees, and scout bees. Bees in the workforce have the duty of finding food sources and communicating their whereabouts to other bees by performing a waggle dance. Onlooker bees then choose the most promising food sources based on the information they received from the employed bees. Scout bees search for new food sources when the current ones are depleted. ABC has proven effective in tackling multiple optimization problems [24], one of which is the initialization of neural network weights [12]. It has shown promising results in improving the performance of deep neural networks and reducing the effect of suboptimal solutions caused by parameter initialization [25]. In addition, the ABC algorithm is a straightforward approach that causes tuning only a few parameters and is simple to implement. As a result, it can be considered a dependable and effective substitute for backpropagation in the training of neural networks.

This paper presents a model called LLC-QE based on deep Q-learning and ensemble learning, combined with the ABC algorithm for weight initialization. Classification is considered a guessing game using a Markov decision process within an RL framework. The state of the environment is represented by a CT image of the patient, and the agent is a deep neural network comprising several parallel convolutional feature extractors. To start the game, the investigation revolves around employing the ABC algorithm in LLC-QE. This algorithm targets the discovery of weight initializations for CNNs and feed-forward networks within the backpropagation algorithm. The agent then decides whether the patient is healthy or ill, and the decision is rewarded with correct decisions receiving positive rewards and incorrect ones receiving negative rewards. In order to address the dataset imbalance, a greater absolute value of the reward is given to the minority class. The aim of the agent is to maximize its cumulative rewards throughout the sequential decision-making process, which involves classifying the samples with the highest possible accuracy. The performance of the LLC-QE model is evaluated on the widely used LIDC-IDRI dataset, and the results show its superiority over other approaches that rely on random weight initialization.

The article is structured in this manner: Section II gives a broad summary of various techniques employed in examining lung nodules. Section III delves deep into the methodology we suggest. Section IV outlines the dataset used for the research and showcases the experimental outcomes. Lastly, Section V

wraps up the discussion and proposes potential avenues for further study.

II. RELATED WORK

CAD is a popular technique for detecting pulmonary nodules in medical images [26]. Conventional methods usually require manually creating features, such as setting pixel thresholds, grouping voxels, and using morphological characteristics. However, these methods are often limited by their ability to detect nodules accurately and to distinguish them from false positives [27]. Tan et al. [28] created a CAD system based on CNN that uses a nodule segmentation technique to detect nodule clusters' central positions in the detection phase. The method integrates computed divergence features with nodule and vessel enhancement filters. In the classification stage, distinctive features that are invariant and defined on a gauge coordinate system are employed to distinguish genuine nodules from certain types of blood vessels that can result in inaccurate positive identifications. Another approach to CAD system development is to merge two or more existing CAD sub-systems to improve accuracy. Traverso and colleagues [29] developed a CAD system that operates through the web and cloud by merging two separate CAD sub-systems: the Channeler Ant Model and the Voxel-Based Neural Approach (VBNA). Both algorithms share a starting point, which involves utilizing a three-dimensional (3D) region-growing segmentation method to obtain the parenchymal volume while simultaneously eliminating the trachea and dividing the two lungs. The Channeler Ant Model utilizes an ant-colony optimization algorithm to identify nodule candidates, while the VBNA uses a multi-layer perceptron neural network to classify them.

In recent times, the field of computer vision has undergone a revolution with the emergence of deep learning, especially CNNs. CNNs have demonstrated remarkable performance in extracting pertinent features from images that can be employed for tasks, such as object detection and classification [30, 31]. This has led to significant advances in fields ranging from medical imaging to autonomous driving. One of the key advantages of CNNs is their ability to learn features in an end-to-end method with no hand-crafted features or feature engineering. This means that CNNs can learn to recognize complex patterns and features in images, such as edges, corners, and textures, by processing the raw pixel values directly. This has led to significant improvements in image classification accuracy on benchmark datasets, such as ImageNet, where CNNs have achieved human-level performance. Another important advantage of CNNs is their ability to generalize to new tasks and datasets. Transfer learning refers to a methodology that enables pre-trained CNNs to apply to new datasets or tasks. This can be done by fine-tuning the model on the new data or by employing the network as a fixed feature extractor. This has been effective for a wide range of tasks, from medical image analysis to natural language processing. There are several well-known frameworks for object detection that utilize CNNs, such as Faster R-CNN [32], SSD [33], and R-FCN [34]. These frameworks use a combination of CNNs and additional modules to generate candidate bounding boxes for objects in an image, which can then be classified and refined to produce the

final object detection output. One of the key advantages of these frameworks is that they can produce highly accurate object detections in a one-stage manner with no complex post-processing or refinement steps.

With the increasing prevalence of deep learning, a growing number of researchers in the field of medical imaging are currently focusing on integrating deep learning into their investigations [35]. For example, many recently proposed CAD systems for identifying pulmonary nodules utilize CNNs to achieve fast and accurate diagnoses. A survey discussed in [36] reveals that multiple CAD systems have emerged to detect nodules comprehensively. ZNET leverages CNNs for detecting candidates and reducing false positives. The input slices are cropped to dimensions of 512×512 , and a U-Net model applies to each axial slice to create a probability map, which is utilized for identifying candidates. Subsequently, a threshold is used to obtain potential nodule regions, which is established with the objective of identifying the maximum possible number of nodules, based on the validation subset. Afterward, a 4-neighborhood kernel is applied for morphological erosion, which aids in eliminating the partial volume effects. A connected component analysis is utilized to group the candidates together, and the coordinates of the candidates are determined based on the centroid of the components. To decrease the occurrence of false positives, ZNET uses wide residual networks [37] and captures $64 * 64$ image sections from the axial, sagittal, and coronal perspectives for each candidate, which are subsequently fed into the networks for independent processing. JianPeiCAD uses a rule-based screening at multiple scales to obtain potential nodules. To decrease false-positive results, a 3D CNN is employed with broad channels and is trained using data augmentation techniques. MOT_M5Lv1 [38] utilizes a technique called 3D region growing to obtain the lung volume, along with specific steps for excluding the trachea and separating the lungs. The

algorithm for detecting candidates is derived from the approach presented by Messay et al. [39], which uses morphological processing and multiple gray-level thresholding to segment nodules. The elimination of false positives is carried out through the calculation of 15 features, which include geometrical and intensity features, and then classification is accomplished by utilizing feedforward neural networks. Resnet [40] suggests a framework for nodule detection using a 3D CNN. This framework screens candidates with a fully convolutional network and selects locations with high probabilities as candidates. To decrease the number of false positives, the recommendation is to incorporate multi-level contextual details surrounding pulmonary nodules by merging a collection of 3D CNNs with distinct receptive field sizes. This approach enables better differentiation of nodules from their challenging imitators. M5LCAD [41] uses ant colonies to segment the lung structures and performs a repetitive process of applying threshold values on the pheromone maps to obtain a set of possible candidates. To reduce the number of false positives, candidates are classified using a feedforward neural network based on a collection of 13 features, which encompass attributes related to spatial positioning, intensity values, and shape characteristics.

III. PROPOSED METHOD

A deep learning framework is being used for binary classification, as shown in Fig. 1. The CT image is taken as input and processed by three CNN feature extractors simultaneously. These extractors individually create a feature vector from the image, and the resulting vectors are merged and passed through fully connected layers, with the last being a Softmax layer that makes the final decision. Using an ensemble of extractors enables the model to generate multi-scale features, which enhance its capabilities and yield better outcomes than a single deep network.

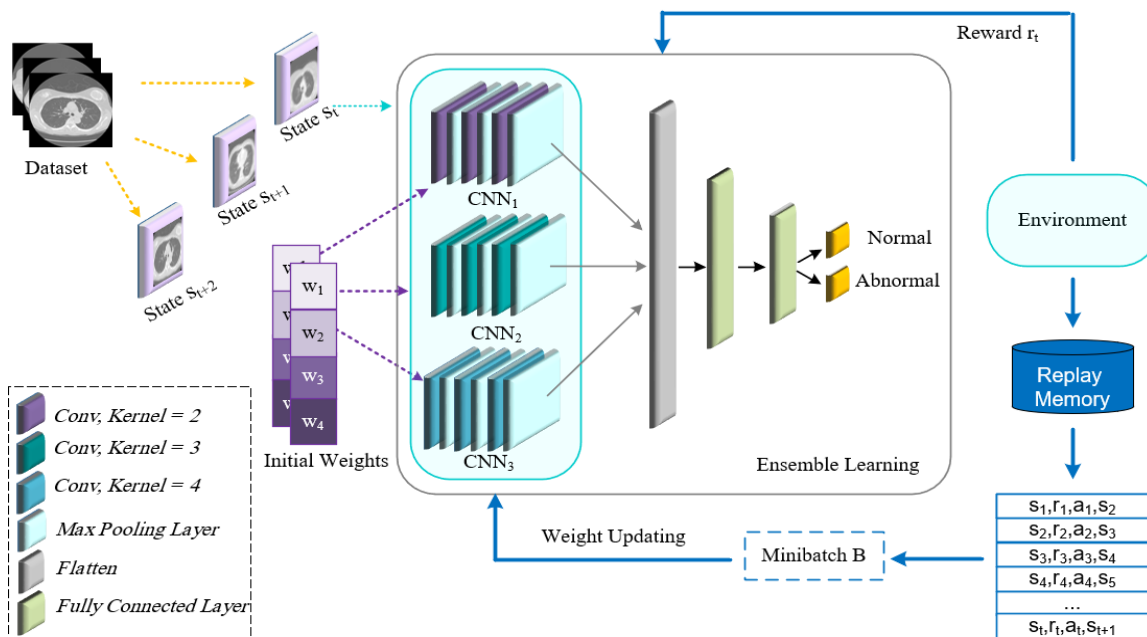


Fig. 1. The LLC-QE model.

The study utilized a CNN network architecture comprised of five convolution layers in two dimensions. The number of filters used in these layers was 128, 64, 32, 16, and 8. Each convolution layer has a kernel size of 2, stride of 3, and padding of 4 for both dimensions. The max-pooling layer has dimensions of 2×2 , and there are three fully connected layers with hidden layer sizes of 128, 64, and 32. In order to avoid overfitting, early stopping and dropouts with a probability of 0.4 are utilized. The batch size for all experiments is 64, and the images in the dataset are gray scale, with light intensities mapped to the range [0,1].

A. Training

The training phase consists of two distinct and sequential steps: ABC pre-training is performed, and deep Q-network training is carried out. The ABC pre-trained weights are used to initialize the deep Q-network training.

1) *ABC pre-training*: The process helps established the network's initial values, increasing the probability of quicker convergence and reducing the chances of getting stuck in local optima. Initially, the weights of the CNN and feedforward layers are transformed into one unified vector, illustrated in Fig. 2. After that, the parameters of each convolutional layer and feedforward layer are compressed and combined into a single vector. Each potential solution for the flattened and concatenated vector is considered a food source in the ABC algorithm. The quality of a solution is evaluated by:

$$Fitness = \frac{1}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (1)$$

The formula evaluates the performance of the algorithm using N training images in the dataset. It considers the actual label y_i and the predicted label \hat{y}_i of the i -th data.

2) *Deep Q-network training*: Every CT image in the training set represents an environmental condition, while the system acts as the operative that performs a series of identifications across all CT pictures. When the operative determines the category tag of the CT picture, it is enacting a step: the picture observed at the t -th instance is the condition s_t , and the identification made is a_t . Consequently, the environment grants a benefit, r_t , to direct the operative. Reward figures are allocated in a manner where identifying an example from the dominant category earns a lesser absolute figure compared to the less common category. The reward function is:

$$r_t(s_t, a_t, y_t) = \begin{cases} +1, & a_t = y_t \text{ and } s_t \in D_S \\ -1, & a_t \neq y_t \text{ and } s_t \in D_S \\ \lambda, & a_t = y_t \text{ and } s_t \in D_H \\ -\lambda, & a_t \neq y_t \text{ and } s_t \in D_H \end{cases} \quad (2)$$

where D_S and D_H denote the less frequent and more prevalent classes, respectively. Properly or improperly categorizing an instance from the dominant class results in a reward of $+\lambda$ or $-\lambda$, with $0 < \lambda < 1$.

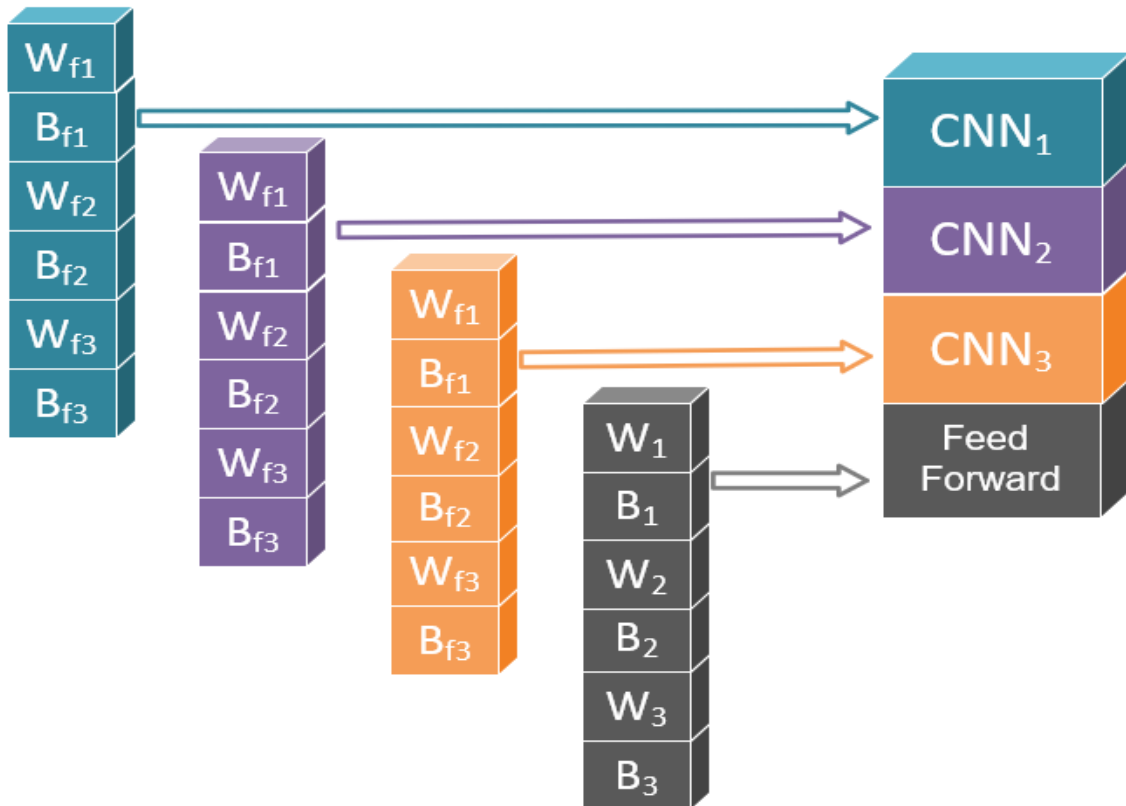


Fig. 2. The weights and biases of the neural network, starting from the initial convolutional layer up to the fully connected layers, are arranged and represented as individual elements in a large vector.

IV. EMPIRICAL EVALUATION

A. Dataset

The Lung Image Database Consortium picture archive (LIDC-IDRI) [42] was established by the Foundation for the National Institutes of Health (FNIH) in collaboration with the Food and Drug Administration (FDA). This collection features a chest CT scan paired with an XML file, which records the annotations made by four radiologists, across 1,018 CT scans from 1,010 listed patients. The marking process comprises two phases aiming to pinpoint every nodule in the CT scans with the utmost precision. During the initial phase, termed the blinded-read phase, each radiologist individually reviewed the scans and identified lesions, noting them as “nodule <3 mm” or “non-nodule ≥3 mm”. In the subsequent unblinded-read phase, each radiologist went over their annotations individually, while being conscious of the undisclosed annotations made by their peers. Within the collection, 7,371 lesions were identified as nodules by at least one radiologist. Out of these, 2,669 lesions were labeled as “nodule 3 mm” by at least one out of the four radiologists, with 928 being agreed upon by all. The 2,669 identified lesions were also provided with detailed nodule characteristics and defined outlines.

B. Experimental Results

Under prior experiments, k-fold cross-validation was employed throughout the study. In pursuit of this aim, the dataset is partitioned into k segments, assigning one for testing while employing the remaining for training. This process is reiterated k times, ensuring each datum is used once for testing and once for training. The resulting cross-validation statistical

outcomes encompass metrics such as minimum, median, maximum mean, and standard deviation. However, mean values are used for comparative analysis.

The proposed approach is compared with ten state-of-the-art systems, including 3D-CNNs [43], ESB-ALL [44], Ali et al. [45], MGI-CNN [46], ODN [47], Xie et al. [8], WOA_AP [48], MEMCAP [4], LungNet-DL [49], and MetaCNN-LC [50]. In addition, comparing the proposed model with three primary methods unveils the impact of the ABC and RL components on the model’s performance. The CNN + random weight method is a model that employs only the CNN network without the ABC algorithm and Reinforcement learning, while the CNN + ABC and CNN + RL models apply ABC and RL, respectively. The model performance on the LIDC dataset with the previously specified criteria is given in Table I and II. Achieving an Accuracy of 92.90%, a Recall of 92.00%, a Precision of 87.70%, an F-measure of 89.80%, a Specificity of 93.40%, and a G-means of 92.70%, the model shows notable distinctions from other deep models. LungNet-DL and MetaCNN-LC are the top two models after the algorithm, with 30% and 40% errors compared to LLC-QE + ABC, respectively. The proposed model reduces the error by over 60% compared with other deep algorithms. Comparing the LLC-QE + ABC model with the LLC + ABC and LLC + ABC models suggests that ABC and RL gimmicks have effectively reduced error by over 52% and 47%, respectively. The worst base model, LLC-QE + random weight, has been improved by approximately 67% by the proposed model.

TABLE I. THE RESULTS OF ACCURACY, RECALL, AND PRECISION FOR THE PROPOSED MODEL AND OTHER ALGORITHMS

Method	Accuracy					Recall					Precision				
	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.
3D-CNNs [43]	0.752	0.789	0.830	0.793	0.029	0.673	0.748	0.804	0.751	0.052	0.621	0.669	0.723	0.672	0.037
ESB-ALL [44]	0.767	0.774	0.849	0.789	0.035	0.664	0.729	0.822	0.736	0.059	0.628	0.651	0.752	0.671	0.048
Ali et al. [45]	0.774	0.814	0.836	0.809	0.022	0.692	0.757	0.766	0.744	0.031	0.655	0.707	0.752	0.705	0.035
MGI-CNN [46]	0.789	0.814	0.858	0.819	0.026	0.738	0.776	0.86	0.787	0.048	0.669	0.711	0.754	0.709	0.033
ODNN [47]	0.789	0.821	0.862	0.821	0.027	0.729	0.766	0.879	0.781	0.058	0.672	0.718	0.752	0.714	0.029
Xie et al. [8]	0.792	0.821	0.871	0.825	0.032	0.748	0.794	0.822	0.785	0.029	0.672	0.719	0.817	0.723	0.059
WOA_AP [48]	0.840	0.865	0.896	0.867	0.023	0.794	0.832	0.897	0.835	0.044	0.746	0.802	0.814	0.785	0.032
MEMCAP [4]	0.858	0.877	0.931	0.887	0.028	0.822	0.860	0.935	0.869	0.044	0.754	0.804	0.870	0.809	0.042
LungNet-DL [49]	0.871	0.886	0.930	0.892	0.036	0.842	0.862	0.923	0.869	0.012	0.776	0.816	0.876	0.824	0.123
MetaCNN-LC [50]	0.885	0.905	0.925	0.902	0.016	0.863	0.875	0.931	0.882	0.026	0.792	0.834	0.887	0.840	0.031
LLC-QE + random weight	0.755	0.792	0.805	0.783	0.023	0.766	0.813	0.841	0.809	0.03	0.603	0.654	0.669	0.641	0.029
LLC + ABC	0.818	0.849	0.881	0.849	0.023	0.785	0.850	0.897	0.843	0.044	0.706	0.744	0.78	0.743	0.026
LLC-QE	0.83	0.865	0.906	0.865	0.027	0.785	0.869	0.897	0.854	0.044	0.73	0.762	0.847	0.77	0.045
LLC-QE + ABC	0.915	0.931	0.943	0.929	0.011	0.888	0.916	0.944	0.920	0.022	0.860	0.883	0.894	0.877	0.015

TABLE II. THE RESULTS OF F-MEASURE, SPECIFICITY, AND G-MEANS FOR THE PROPOSED MODEL AND OTHER ALGORITHMS

Method	F-measure					Specificity					G-means				
	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.
3D-CNNs [43]	0.646	0.702	0.761	0.709	0.043	0.791	0.815	0.844	0.814	0.019	0.730	0.776	0.824	0.782	0.035
ESB-ALL [44]	0.657	0.686	0.785	0.701	0.049	0.773	0.820	0.863	0.816	0.032	0.738	0.765	0.842	0.775	0.040
Ali et al. [45]	0.673	0.733	0.759	0.724	0.032	0.815	0.839	0.872	0.842	0.020	0.751	0.799	0.817	0.791	0.025
MGI-CNN [46]	0.702	0.733	0.804	0.746	0.039	0.815	0.844	0.858	0.836	0.018	0.776	0.799	0.859	0.811	0.032
ODNN [47]	0.699	0.742	0.811	0.746	0.041	0.820	0.844	0.853	0.842	0.013	0.773	0.806	0.866	0.811	0.034
Xie et al. [8]	0.708	0.742	0.805	0.752	0.038	0.806	0.848	0.910	0.845	0.041	0.781	0.806	0.850	0.814	0.028
WOA_APSO [48]	0.769	0.798	0.853	0.809	0.034	0.863	0.896	0.900	0.884	0.019	0.828	0.847	0.896	0.859	0.028
MEMCAP [4]	0.804	0.822	0.901	0.838	0.040	0.858	0.896	0.929	0.896	0.025	0.858	0.868	0.932	0.882	0.031
LungNet-DL [49]	0.825	0.841	0.894	0.864	0.022	0.862	0.905	0.923	0.913	0.035	0.862	0.885	0.905	0.892	0.014
MetaCNN-LC [50]	0.842	0.863	0.906	0.896	0.026	0.876	0.914	0.935	0.926	0.032	0.872	0.896	0.914	0.906	0.009
LLC-QE + random weight	0.683	0.725	0.742	0.715	0.029	0.735	0.782	0.791	0.770	0.023	0.762	0.797	0.811	0.789	0.024
LLC + ABC	0.743	0.791	0.834	0.789	0.033	0.834	0.848	0.872	0.852	0.014	0.809	0.849	0.884	0.847	0.028
LLC-QE	0.757	0.812	0.863	0.810	0.038	0.848	0.863	0.919	0.870	0.028	0.818	0.866	0.899	0.862	0.030
LLC-QE + ABC	0.876	0.899	0.918	0.898	0.017	0.924	0.938	0.943	0.934	0.008	0.908	0.927	0.943	0.927	0.014

The aim is to carry out an additional experiment to assess the influence of employing distinct algorithms for initializing the model parameters. To achieve this aim, in order to maintain a fair assessment, all components of the model will remain unchanged—encompassing reinforcement learning and the CNN structure—with alterations limited solely to the initialization

algorithm. Substitution of the algorithmic instructor will involve five established conventional algorithms, including GDM [51], GDA [52], GDMA [53], OSS [54], and BR [55], and four metaheuristic algorithms, including GWO [56], BA [57], COA [58] and WOA [59]. The ABC algorithm used in the model outperforms all other meta-heuristic algorithms (Table III and IV).

TABLE III. THE RESULTS OF ACCURACY, RECALL, AND PRECISION FOR THE CONVENTIONAL AND METAHEURISTIC ALGORITHMS

Method	Accuracy					Recall					Precision				
	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.
LLC-QE + GDM	0.805	0.865	0.906	0.864	0.039	0.748	0.794	0.888	0.806	0.059	0.696	0.802	0.841	0.793	0.058
LLC-QE + GDA	0.824	0.881	0.899	0.872	0.029	0.785	0.841	0.879	0.841	0.035	0.718	0.800	0.832	0.793	0.046
LLC-QE + GDMA	0.827	0.849	0.865	0.848	0.014	0.776	0.785	0.841	0.798	0.028	0.728	0.764	0.79	0.763	0.023
LLC-QE + OSS	0.849	0.871	0.884	0.867	0.015	0.748	0.813	0.869	0.806	0.056	0.788	0.806	0.813	0.801	0.011
LLC-QE + BR	0.843	0.862	0.884	0.863	0.016	0.757	0.804	0.869	0.806	0.050	0.771	0.789	0.802	0.791	0.013
LLC-QE + GWO	0.858	0.865	0.896	0.869	0.016	0.776	0.813	0.841	0.807	0.027	0.786	0.798	0.849	0.805	0.025
LLC-QE + BAT	0.852	0.868	0.877	0.865	0.010	0.766	0.785	0.832	0.794	0.031	0.778	0.804	0.828	0.804	0.019
LLC-QE + COA	0.821	0.877	0.881	0.864	0.025	0.738	0.832	0.841	0.811	0.044	0.731	0.804	0.811	0.79	0.034
LLC-QE + WOA	0.833	0.881	0.899	0.874	0.026	0.841	0.841	0.879	0.852	0.017	0.714	0.811	0.832	0.792	0.049

TABLE IV. THE RESULTS OF F-MEASURE, SPECIFICITY, AND G-MEANS FOR THE CONVENTIONAL AND METAHEURISTIC ALGORITHMS

Method	F-measure					Specificity					G-means				
	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.	min	median	max	mean	std.dev.
LLC-QE + GDM	0.721	0.798	0.864	0.799	0.056	0.834	0.900	0.915	0.893	0.034	0.79	0.845	0.901	0.848	0.043
LLC-QE + GDA	0.750	0.829	0.855	0.816	0.040	0.844	0.891	0.910	0.888	0.027	0.814	0.875	0.894	0.864	0.03
LLC-QE + GDMA	0.751	0.783	0.807	0.780	0.020	0.853	0.877	0.896	0.874	0.016	0.814	0.834	0.859	0.835	0.016
LLC-QE + OSS	0.769	0.809	0.831	0.803	0.028	0.882	0.900	0.910	0.898	0.010	0.82	0.855	0.875	0.85	0.026
LLC-QE + BR	0.764	0.796	0.834	0.797	0.028	0.886	0.891	0.905	0.892	0.008	0.819	0.846	0.88	0.847	0.025
LLC-QE + GWO	0.787	0.804	0.845	0.806	0.024	0.886	0.896	0.924	0.900	0.014	0.836	0.853	0.882	0.853	0.018
LLC-QE + BAT	0.781	0.796	0.818	0.799	0.016	0.886	0.905	0.919	0.901	0.013	0.833	0.839	0.863	0.846	0.015
LLC-QE + COA	0.734	0.820	0.826	0.800	0.038	0.863	0.896	0.900	0.891	0.016	0.798	0.865	0.87	0.85	0.03
LLC-QE + WOA	0.772	0.826	0.855	0.821	0.033	0.829	0.900	0.910	0.885	0.034	0.835	0.87	0.894	0.869	0.023

1) *Impact of the reward function:* The rewards for accurate and erroneous categorizations are given to the predominant and less frequent classes as ± 1 and $\pm \lambda$, respectively. The λ value is influenced by the ratio of dominant to fewer common examples, and it is expected that as this ratio rises, the ideal λ value will drop. To explore the influence of λ , we evaluated the suggested model's effectiveness across various λ values, which spanned from 0 to 1, at 0.1 intervals, while keeping the rewards for the dominant class unaltered. These outcomes are depicted in Fig. 3. When λ is zero, the dominant class's influence is minimal, and at $\lambda = 1$, both classes have equivalent influences. Fig. 3 reveals that

the model's optimal performance is achieved when λ is 0.4, across all evaluated metrics. This suggests that the best λ value is neither zero nor one, but falls somewhere between these extremes. It is crucial to highlight that, while it is essential to reduce the dominant class's influence by tweaking λ , setting it excessively low might degrade the model's overall effectiveness. The findings indicate that selecting an appropriate λ value profoundly affects the LLC-QE model's efficiency. The best λ value is influenced by the respective quantities of dominant and less frequent examples, making it crucial to determine it prudently for optimal outcomes.

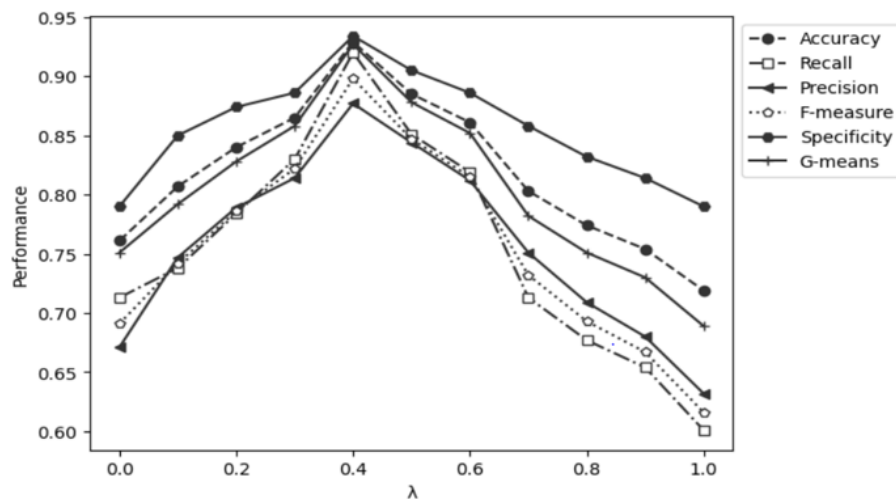


Fig. 3. LLC-QE model performance metrics plotted against the value of λ in the reward function.

2) *Impact of the number of CNNs*: The LLC-QE model uses a group of CNNs to derive feature vectors from input images simultaneously. The number of CNN feature extractors can significantly impact the model's overall performance. Using too few CNNs could cause insufficient feature extraction, while using too many CNNs could lead to overfitting or redundant information extraction, both of which may negatively impact the model's performance. The performance of the LLC-QE model was assessed by altering the count of CNN feature extractors across the range of one to seven. This variation is aimed at identifying the optimal number of extractors. The findings indicated optimal performance was achieved when three CNNs were used, as shown in Fig. 4. The model's performance decreased as the number of CNNs increased, with performing six and seven CNNs being worse than that of a single CNN. The optimal number of CNNs was determined based on performance metrics.

3) *Impact of the loss function*: Classification problems caused by imbalanced datasets can also address using conventional methods, such as altering the loss function and using data augmentation. However, their effectiveness depends highly on the specific problem being addressed.

Meanwhile, the loss function plays a more significant role, as it can give more prominence to the minority class. To study the inefficiency of the loss functions on the training ANN, the selection encompassed five functions, including Weighted Cross-Entropy (WCE) [60], Balanced Cross-Entropy (BCE) [61], Dice Loss (DL) [62], Tversky Loss (TL) [63], and Focal Loss (FL) [64]. WCE and BCE both assign weights to positive and negative samples. FL, suited for imbalanced data, outperforms the other loss function (Table V) but is still inferior to the RL used in the model.

4) *Impact of pre-trained models*: Comparing the performance of the CNN ensemble model with that of alternative pre-trained feature extraction models involved replacing the model with transfer learning counterparts, such as AlexNet [65], GoogleNet [66], ResNet [67], DenseNet [68], and MobileNet [69]. Limiting the training solely to the feedforward network, superior performance is exhibited by the model's ensemble of CNNs. This ensemble, trained from the ground up, surpasses the performance of pre-trained networks (AlexNet, GoogleNet, ResNet, DenseNet, and MobileNet), as demonstrated in Table VI. The reason behind this is that the ensemble of CNNs is more capable of extracting discriminative features specific to cancer diagnosis.

TABLE V. PERFORMANCE OF THE PROPOSED MODEL FOR DIFFERENT LOSS FUNCTIONS

Method	Accuracy	Sensitivity	Precision	F-measure	Specificity	G-mean
WCE	0.830	0.885	0.790	0.803	0.778	0.845
BCE	0.822	0.821	0.783	0.745	0.824	0.822
DL	0.816	0.795	0.769	0.711	0.837	0.811
TL	0.825	0.834	0.78	0.747	0.816	0.827
FL	0.861	0.889	0.826	0.819	0.833	0.868

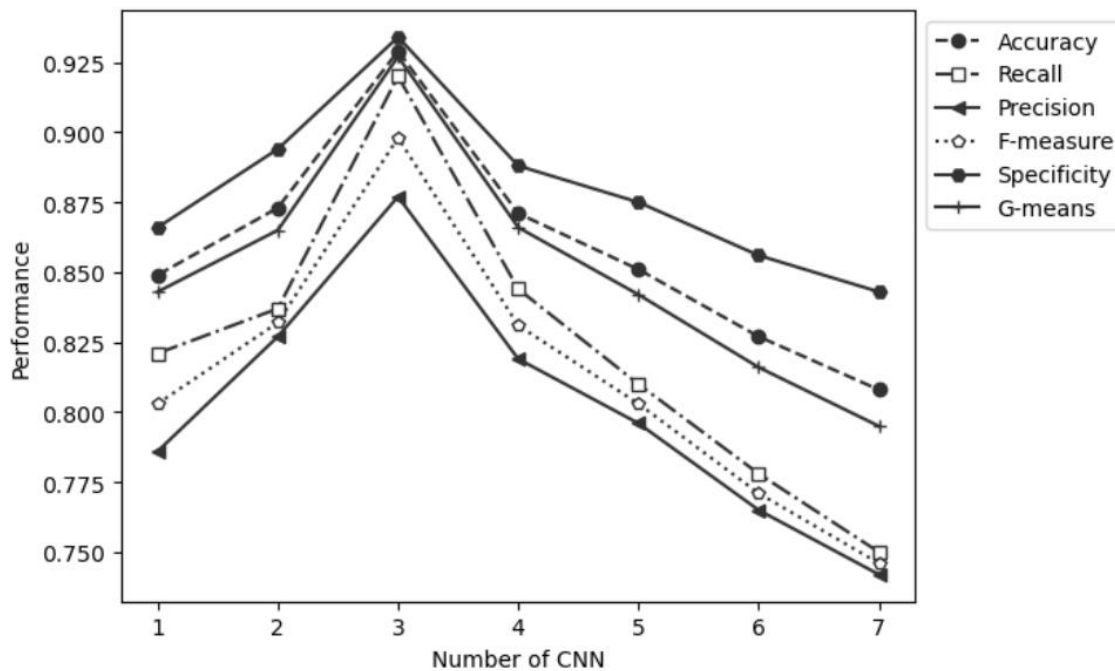


Fig. 4. The performance metrics plotted vs the number of convolutional feature extractors working in the ensemble.

TABLE VI. PERFORMANCE OF THE PROPOSED MODEL FOR DIFFERENT PRE-TRAINED MODELS

Method	Accuracy	Sensitivity	Precision	F-measure	Specificity	G-mean
AlexNet	0.773	0.836	0.719	0.724	0.715	0.790
GoogleNet	0.789	0.768	0.737	0.676	0.811	0.784
ResNet	0.764	0.794	0.709	0.683	0.736	0.772
DenseNet	0.753	0.759	0.696	0.651	0.748	0.755
MobileNet	0.784	0.806	0.731	0.703	0.762	0.789

V. CONCLUSION

This article proposes a novel deep-learning approach that uses RL and evolutionary computation to classify lung cancer in CT images. To avoid the model getting trapped in local optima, the network weights are first initialized using the evolutionary ABC algorithm. The network architecture comprises an ensemble of CNNs that extract features in parallel and then concatenate them for downstream classification. The model uses RL to address the dataset imbalance. The proposed LLC-QE model achieves excellent results compared to other deep learning models and pre-trained transfer learning models when trained on the LIDC-IDRI dataset. The optimal value for the reward function and the optimal number of CNN feature extractors in the ensemble are determined through experiments on the study dataset. Separate ablation studies, excluding ABC pre-training and RL, confirm the positive incremental impact of these components on model performance. Notably, the ABC algorithm and RL outperform various meta-heuristic initialization algorithms and loss functions.

Future work aims to develop deep learning segmentation methods that can detect not only the presence of cancer but also the location and extent of the disease on CT images, which may be useful for prognostication and therapeutic monitoring. One area of research that holds particular promise is the use of multi-modal imaging data, which can provide a more comprehensive view of the tumor and its surroundings. For example, combining CT with MRI data could allow for more accurate identification of the tumor boundary and help differentiate between different cancers.

FUNDING

This work was supported by Science and Technology Research Project of Jiangxi Provincial Department of Education (GJJ2202030), and Natural Science Research Project of Nanchang Normal University (22XJZR02).

REFERENCES

[1] S. Danaei et al., "Myocarditis Diagnosis: A Method using Mutual Learning-Based ABC and Reinforcement Learning," in 2022 IEEE 22nd International Symposium on Computational Intelligence and Informatics and 8th IEEE International Conference on Recent Achievements in Mechatronics, Automation, Computer Science and Robotics (CINTI-MACRo), 2022: IEEE, pp. 000265-000270.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA: a cancer journal for clinicians*, vol. 66, no. 1, pp. 7-30, 2016.

[3] H. Zareiamand, A. Darroudi, I. Mohammadi, S. V. Moravvej, S. Danaei, and R. Alizadehsani, "Cardiac Magnetic Resonance Imaging (CMRI) Applications in Patients with Chest Pain in the Emergency Department: A Narrative Review," *Diagnostics*, vol. 13, no. 16, p. 2667, 2023.

[4] A. Mobiny et al., "Memory-augmented capsule network for adaptable lung nodule classification," *IEEE Transactions on Medical Imaging*, 2021.

[5] L. Hong et al., "GAN-LSTM-3D: An efficient method for lung tumour 3D reconstruction enhanced by attention-based LSTM," *CAAI Transactions on Intelligence Technology*, 2023.

[6] S. V. Moravvej et al., "RLMD-PA: A reinforcement learning-based myocarditis diagnosis combined with a population-based algorithm for pretraining weights," *Contrast Media & Molecular Imaging*, vol. 2022, 2022.

[7] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE transactions on medical imaging*, vol. 25, no. 4, pp. 385-405, 2006.

[8] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109-119, 2019.

[9] J. Ning, H. Zhao, L. Lan, P. Sun, and Y. Feng, "A computer-aided detection system for the detection of lung nodules based on 3D-ResNet," *Applied Sciences*, vol. 9, no. 24, p. 5544, 2019.

[10] M. A. Wiering, H. Van Hasselt, A.-D. Pietersma, and L. Schomaker, "Reinforcement learning algorithms for solving classification problems," 2011: IEEE, pp. 91-96.

[11] N. Tajbakhsh and K. Suzuki, "Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs," *Pattern recognition*, vol. 63, pp. 476-486, 2017.

[12] S. V. Moravvej, S. J. Mousavirad, D. Oliva, and F. Mohammadi, "A Novel Plagiarism Detection Approach Combining BERT-based Word Embedding, Attention-based LSTMs and an Improved Differential Evolution Algorithm," *arXiv preprint arXiv:2305.02374*, 2023.

[13] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning," in *International conference on intelligent computing*, 2005: Springer, pp. 878-887.

[14] I. Mani and I. Zhang, "kNN approach to unbalanced data distributions: a case study involving information extraction," in *Proceedings of workshop on learning from imbalanced datasets*, 2003, vol. 126: ICML United States.

[15] S. V. Moravvej, S. J. Mousavirad, M. H. Moghadam, and M. Saadatmand, "An LSTM-based plagiarism detection via attention mechanism and a population-based approach for pre-training parameters with imbalanced classes," in *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part III 28, 2021*: Springer, pp. 690-701.

[16] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26-38, 2017.

[17] M. A. Wiering, H. Van Hasselt, A.-D. Pietersma, and L. Schomaker, "Reinforcement learning algorithms for solving classification problems," in *2011 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, 2011: IEEE, pp. 91-96.

[18] T. Zhang, M. Huang, and L. Zhao, "Learning structured representation for text classification via reinforcement learning," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[19] D. Liu and T. Jiang, "Deep reinforcement learning for surgical gesture segmentation and classification," in *International conference on medical image computing and computer-assisted intervention*, 2018: Springer, pp. 247-255.

[20] D. Zhao, Y. Chen, and L. Lv, "Deep reinforcement learning with visual attention for vehicle classification," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 4, pp. 356-367, 2016.

[21] J. Janisch, T. Pevný, and V. Lisý, "Classification with costly features using deep reinforcement learning," in *Proceedings of the AAAI*

- Conference on Artificial Intelligence, 2019, vol. 33, no. 01, pp. 3959-3966.
- [22] L. Abdi and S. Hashemi, "An ensemble pruning approach based on reinforcement learning in presence of multi-class imbalanced data," in Proceedings of the Third International Conference on Soft Computing for Problem Solving, 2014: Springer, pp. 589-600.
- [23] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the artificial bee colony (ABC) algorithm in collaboration with the fog nodes in the Internet of Things three-layer architecture," in 2021 29th Iranian Conference on Electrical Engineering (ICEE), 2021: IEEE, pp. 509-513.
- [24] S. Vakilian, S. V. Moravvej, and A. Fanian, "Using the cuckoo algorithm to optimizing the response time and energy consumption cost of fog nodes by considering collaboration in the fog layer," in 2021 5th International Conference on Internet of Things and Applications (IoT), 2021: IEEE, pp. 1-5.
- [25] S. V. Moravvej, S. J. Mousavirad, D. Oliva, G. Schaefer, and Z. Sobhaninia, "An improved de algorithm to optimise the learning process of a bert-based plagiarism detection model," in 2022 IEEE Congress on Evolutionary Computation (CEC), 2022: IEEE, pp. 1-7.
- [26] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and therapy*, pp. 2015-2022, 2015.
- [27] H. Xie, Y. Zhang, K. Gao, S. Tang, K. Xu, L. Guo, and J. Li, "Robust common visual pattern discovery using graph matching," *Journal of visual communication and image representation*, vol. 24, no. 5, pp. 635-646, 2013.
- [28] M. Tan, R. Deklerck, B. Jansen, M. Bister, and J. Cornelis, "A novel computer-aided lung nodule detection system for CT images," *Medical physics*, vol. 38, no. 10, pp. 5630-5645, 2011.
- [29] A. Traverso, E. L. Torres, M. Fantacci, and P. Cerello, "Computer-aided detection systems to improve lung cancer early diagnosis: state-of-the-art and challenges," in *Journal of Physics: Conference Series*, 2017, vol. 841, no. 1: IOP Publishing, p. 012013.
- [30] T. Liu, W. Xu, P. Spincemaille, A. S. Avestimehr, and Y. Wang, "Accuracy of the morphology enabled dipole inversion (MEDI) algorithm for quantitative susceptibility mapping in MRI," *IEEE transactions on medical imaging*, vol. 31, no. 3, pp. 816-824, 2012.
- [31] H. Xie, K. Gao, Y. Zhang, and J. Li, "Local geometric consistency constraint for image retrieval," in 2011 18th IEEE International Conference on Image Processing, 2011: IEEE, pp. 101-104.
- [32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, 2016: Springer, pp. 21-37.
- [34] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, vol. 29, 2016.
- [35] K.-L. Hua, H.-C. Wang, C.-H. Yeh, W.-H. Cheng, and Y.-C. Lai, "Background extraction using random walk image fusion," *IEEE transactions on cybernetics*, vol. 48, no. 1, pp. 423-435, 2016.
- [36] A. A. A. Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge," *Medical image analysis*, vol. 42, pp. 1-13, 2017.
- [37] S. Zagoruyko and N. Komodakis, "Wide residual networks," *arXiv preprint arXiv:1605.07146*, 2016.
- [38] Q. Dou, H. Chen, Y. Jin, H. Lin, J. Qin, and P.-A. Heng, "Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, 2017: Springer, pp. 630-638.
- [39] T. Messay, R. C. Hardie, and S. K. Rogers, "A new computationally efficient CAD system for pulmonary nodule detection in CT imagery," *Medical image analysis*, vol. 14, no. 3, pp. 390-406, 2010.
- [40] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558-1567, 2016.
- [41] E. Lopez Torres et al., "Large scale validation of the M5L lung CAD on heterogeneous CT datasets," *Medical physics*, vol. 42, no. 4, pp. 1477-1489, 2015.
- [42] L. M. Pehrson, M. B. Nielsen, and C. Ammitzbøl Lauridsen, "Automatic pulmonary nodule detection applying deep learning or machine learning algorithms to the LIDC-IDRI database: a systematic review," *Diagnostics*, vol. 9, no. 1, p. 29, 2019.
- [43] P. Monkam, S. Qi, M. Xu, H. Li, F. Han, Y. Teng, and W. Qian, "Ensemble learning of multiple-view 3D-CNNs model for micro-nodules identification in CT images," *IEEE Access*, vol. 7, pp. 5564-5576, 2018.
- [44] H. Jung, B. Kim, I. Lee, J. Lee, and J. Kang, "Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method," *BMC medical imaging*, vol. 18, no. 1, pp. 1-10, 2018.
- [45] I. Ali et al., "Lung nodule detection via deep reinforcement learning," *Frontiers in oncology*, vol. 8, p. 108, 2018.
- [46] B.-C. Kim, J. S. Yoon, J.-S. Choi, and H.-I. Suk, "Multi-scale gradual integration CNN for false positive reduction in pulmonary nodule detection," *Neural Networks*, vol. 115, pp. 1-10, 2019.
- [47] S. Lakshmanaprabu, S. N. Mohanty, K. Shankar, N. Arunkumar, and G. Ramirez, "Optimal deep learning model for classification of lung cancer on CT images," *Future Generation Computer Systems*, vol. 92, pp. 374-382, 2019.
- [48] S. Vijh, P. Gaurav, and H. M. Pandey, "Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection," *Neural Computing and Applications*, pp. 1-14, 2020.
- [49] R. Tandon, S. Agrawal, R. Raghuvanshi, N. P. S. Rathore, L. Prasad, and V. Jain, "Automatic lung carcinoma identification and classification in CT images using CNN deep learning model," in *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*: Springer, 2022, pp. 143-166.
- [50] T. I. Mohamed, O. N. Oyelade, and A. E. Ezugwu, "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm," *PloS one*, vol. 18, no. 8, p. e0285796, 2023.
- [51] V. Phansalkar and P. Sastry, "Analysis of the back-propagation algorithm with momentum," *IEEE Transactions on Neural Networks*, vol. 5, no. 3, pp. 505-506, 1994.
- [52] M. Hagan, H. Demuth, and M. Beale, "Neural Network Design (PWS, Boston, MA)," *Google Scholar Google Scholar Digital Library Digital Library*, 1996.
- [53] C.-C. Yu and B.-D. Liu, "A backpropagation algorithm with adaptive learning rate and momentum coefficient," in *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)*, 2002, vol. 2: IEEE, pp. 1218-1223.
- [54] R. Battiti, "First-and second-order methods for learning: between steepest descent and Newton's method," *Neural computation*, vol. 4, no. 2, pp. 141-166, 1992.
- [55] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proceedings of international conference on neural networks (ICNN'97)*, 1997, vol. 3: IEEE, pp. 1930-1935.
- [56] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in engineering software*, vol. 69, pp. 46-61, 2014.
- [57] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in *Nature inspired cooperative strategies for optimization (NICSO 2010)*: Springer, 2010, pp. 65-74.
- [58] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World congress on nature & biologically inspired computing (NaBIC)*, 2009: Ieee, pp. 210-214.
- [59] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in engineering software*, vol. 95, pp. 51-67, 2016.

- [60] V. Pihur, S. Datta, and S. Datta, "Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach," *Bioinformatics*, vol. 23, no. 13, pp. 1607-1615, 2007.
- [61] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1395-1403.
- [62] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*: Springer, 2017, pp. 240-248.
- [63] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," in *International workshop on machine learning in medical imaging*, 2017: Springer, pp. 379-387.
- [64] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980-2988.
- [65] M. Z. Alom et al., "The history began from alexnet: A comprehensive survey on deep learning approaches," *arXiv preprint arXiv:1803.01164*, 2018.
- [66] R. Anand, T. Shanthi, M. Nithish, and S. Lakshman, "Face recognition and classification using GoogleNET architecture," in *Soft Computing for Problem Solving: SocProS 2018, Volume 1*, 2020: Springer, pp. 261-269.
- [67] S. Targ, D. Almeida, and K. Lyman, "Resnet in resnet: Generalizing residual architectures," *arXiv preprint arXiv:1603.08029*, 2016.
- [68] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700-4708.
- [69] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

Secure Data Sharing in Smart Homes: An Efficient Approach Based on Local Differential Privacy and Randomized Responses

Amr T. A. Elsayed¹, Almohammady S. Alsharkawy², Mohamed S. Farag³, S. E. Abo-Youssef⁴
Faculty of Science, Al-Azhar University, Cairo, Egypt^{1,2,4}
Obour Heigh Institute for informatics, Cairo, Egypt³

Abstract—Smart homes are smart spaces that contain devices that are connected to each other, collecting information and facilitating users' comfortable living, safety, and energy management features. To improve the quality of individuals' life, smart device companies and service providers are collecting data about user activities, user needs, power consumption, etc.; these data need to be shared with companies with privacy-preserving practices. In this paper, an effective approach of securing data transmission to the service provider is based on local differential privacy (LDP), which enables residents of smart homes to provide statistics on their power usage as disturbances bloom filters. Randomized Aggregatable Privacy-Preserving Ordinal (RAPPOR) is a privacy technique that allows sharing of data and statistics while preserving the privacy of individual users. The proposed approach applies two randomized responses: permanent random response (PRR) and instantaneous random response (IRR), then applies machine learning algorithms for decoding the perturbation bloom filters on the service provider side. The simulation results show that the proposed approach achieves good performance in terms of privacy-preserving, accuracy, recall, and f-measure metrics. The results indicate that, the proposed LDP for smart homes achieved good utility privacy when the value of LDP $\epsilon = 0.95$. The classification accuracy is between 95.4% and 98% for the utilized classification techniques.

Keywords—Smart homes; security; privacy-preserving; differential privacy; RAPPOR; randomized responses

I. INTRODUCTION

As more people seek to automate their homes and improve their quality of life, the popularity of smart homes increases. A smart home is a residence that contains remote-controllable devices, such as smart thermostats, security systems, lighting, and entertainment systems. These devices are internet-connected, allowing homeowners to control them remotely using smartphones or other internet-connected devices [1]. Convenience is the main advantage of a smart home it enables users to control the temperature, lighting, and security of your smart home from anywhere in the world. The ability to turn off lights, adjust the temperature, and view security cameras from your smartphone makes it simple to keep your home comfortable and secure. Smart homes can also reduce your energy costs, smart thermostats can automatically adjust your home's temperature based on your preferences and your presence, saving you money on heating and cooling expenses. Similarly, intelligent lighting systems can turn off lights automatically when no one is in a room, thereby reducing energy consumption [2].

A smart home also provides better protection; with intelligent security systems, you can monitor your house from anywhere and receive warnings if suspicious behavior is detected [3]. You can also lock and unlock doors remotely, allowing you to let guests or service personnel in without being present. Smart homes can also improve your entertainment experience. You can control your television, music, and other entertainment systems from anywhere in your smart home. Even your smart home can be integrated with your voice assistant, making it simple to control your entertainment with voice commands.

Data gathered from smart homes can be used in a variety of ways to improve services across a range of industries. These services such as smart home activity prediction [4], smart healthcare for patient treatment [5], disorder assessment, and smart city pedestrian monitoring [6], energy management. In this context, businesses have discovered the potential of using the data gathered from smart homes to improve their products and services.

However, data collectors must consider the confidentiality of these data. If data is not correctly managed, it could cause significant issues. So, to address these concerns, a new system that maintains both privacy and utility has been proposed. Remote health systems necessitate the collection, disclosure, and utilization of personal health information, which raises grave privacy concerns. For many individuals, the household is their most private environment. A glucometer measuring the blood sugar level, a spirometer tracking the air entering and exiting the lungs, and a sleep monitoring sensor recording the sleep conditions can potentially reveal whether a resident has diabetes, seasonal allergy-induced asthma, or a depressive disorder. Patients are inclined to restrict access to these data to a small group, such as their personal physicians, out of concern for their privacy.

Differential Privacy [7] is a privacy preservation mechanism that has gained popularity. The main idea behind differential privacy is that a user is given plausible deniability by adding random values to their input. This approach provides strong privacy guarantees for users, protecting their data against adversary entities, such as service providers and outsiders. In the centralized differential privacy setting, noise is added to the database and apply a differential privacy aggregation algorithm. RAPPOR [8] is a privacy preservation technology that allows for the sharing of statistics while preserving the privacy of individual users. By using randomized response, RAPPOR ensures that no individual's data is

disclosed to the data collector. This approach has shown great promise, as it allows for the sharing of valuable data while protecting users' privacy.

This paper aims to present an approach for securely transmitting household data to the aggregator, while accounting for the presence of malicious aggregator nodes. To address this concern, we apply LDP to the real-time data collected from residences. Prior to transmission, the data is subjected to a process of privacy preservation. The proposed model utilizes the RAPPOR algorithm to encrypt the data, thereby ensuring that the aggregator cannot ascertain the identity of the householder, thereby preserving anonymity. To achieve the goal of secure data transmission, we propose a three-step approach. Firstly, bloom Filter mechanism is applied to the raw data collected from the residences. Next, the data is privatized using the RAPPOR algorithm to ensure that the identity of the householder remains unknown. Finally, the aggregator employs machine learning algorithms to decode the data into a form that is acceptable and useful.

The proposed model has several advantages over existing approaches. By employing RAPPOR, we are able to ensure that the data is secure and anonymous, thereby preventing malicious aggregator nodes from accessing sensitive information. Moreover, the use of machine learning algorithms by the aggregator allows for efficient decoding of the data, making it more accessible and user-friendly. The remainder of this paper is organized as follows. Section II surveys existing privacy preservation techniques used in smart homes and highlighting their deficiencies. Section III gives the useful background about Local Differential Privacy and RAPPOR. Then our approach is described in Sections IV and introduces the system model. In section V, the performance of the scheme is analyzed from two aspects of security and efficiency.

II. RELATED WORK

In recent years, LDP has emerged as a promising technique for privacy-preserving data analysis in various domains. This section provides an overview of some well-known LDP use cases and privacy-preserving systems that have used LDP. In [8] Google proposed RAPPOR as an LDP-based system for collecting aggregate statistics from users without compromising their individual privacy. It randomizes user responses to a question with a bloom filter and randomized response, allowing the server to compute meaningful statistics about the aggregate responses while ensuring individual privacy. The authors in [9] proposed an approach called a differential privacy-based system to guarantee thorough security for data produced by smart houses. At the aggregator level, they used the Hidden Markov Model (HMM) technique and applied differential privacy to the personal information obtained from smart homes.

In healthcare field the authors in [10] proposed an improved approach based on k-anonymity and differential privacy to enhance privacy protection by mitigating re-identification risks through generalization and suppression techniques. This study [11] concentrates primarily on identifying the security issues that can arise from the use of a large number of Internet of Things (IoT) devices connected to provide a smart home facility in Saudi Arabia. [12] proposed an approach called

LATENT, suggests an intermediate layer in deep learning models that satisfies LDP. LATENT allows a data owner to perturb the data on their device before it reaches an untrusted machine learning service, thereby protecting the privacy of the owner's data. By adding noise to the data in a controlled manner, LATENT ensures that the machine learning model can still provide useful insights while preserving the privacy of the individual data points. In Microsoft, LDP is used to collect data about the time users spend in different applications, which enables the identification of their favorite ones and improves their user experience [13]. This approach still preserves user privacy while providing valuable insights for application developers. LDP has also been used to reduce potential privacy leakage in deep learning models.

Differential privacy is a privacy-preserving technique that has been extensively researched for various applications in computer science. One of the most popular applications is in recommendation systems, where differential privacy is used to protect the privacy of user preferences and behavior while still allowing the system to make accurate recommendations [14], [15]. Data mining is another field that benefits from differential privacy, as it allows for the analysis of sensitive data without revealing individual records [16]. Differential privacy is also used in crowd-sourcing [17]. In network measurements, differential privacy is used to ensure that the privacy of individuals' network traffic data is protected while still allowing for useful aggregate network measurements to be obtained [18]. In intelligent transportation systems, differential privacy is used to protect the privacy of users and their data [19].

These approaches have a few disadvantages or limitations compared to our approach, they use a trusted third party to collect data from users, apply some algorithms, and take some privacy-preserving data analysis by adding "noise". This leads to a reduction in the accuracy of data analysis and inference. The noise introduced to protect privacy may make it challenging to obtain precise information or draw accurate conclusions from the collected data. Our solution uses differential privacy at the data source, thereby providing greater privacy. In addition to the use of LDP, researchers have also put forth schemes that employ data masking techniques [20–24]. These approaches involve masking the data submitted by users with a specific masking value, ensuring that other entities cannot access the actual value unless they possess knowledge of the masking value. By incorporating data masking alongside LDP, these schemes offer an extra layer of privacy protection and enhance the security of sensitive information in the context of data sharing and analysis. In each of these applications, differential privacy should provide a way to perform valuable computations on sensitive data while ensuring that the privacy of individual users is protected. By adding controlled noise to the data, differential privacy makes it difficult for attackers to identify any specific individual in the dataset, while still allowing for meaningful analysis and insights to be drawn from the data.

III. PRELIMINARIES

This section provides background information on LDP and the randomized response approach. It also discusses RAPPOR, which is a method for implementing the randomized response

strategy. In addition to this, it investigates the machine learning methods that have been implemented, such as K-nearest neighbours (KNN), Support vector machines (SVMs) and XGBoost. In the final part of this discussion, we will examine the performance and assessment measures that are utilized in this paper to evaluate the performance of the proposed scheme.

A. Local Differential Privacy (LDP)

LDP is a privacy-preserving technique that aims to protect the privacy of individual data contributors while enabling statistical analysis and inference on the aggregated data. Unlike other privacy-preserving methods that rely on centralizing and anonymizing data, LDP allows data contributors to locally perturb their data before sharing it.

Definition: A randomized algorithm T satisfies the ϵ -local differential privacy where $\epsilon > 0$ if for all pairs of the client's values a and b and for all $S \subseteq \text{range}(T)$:

$$Pr[T(a) \in S] \leq e^\epsilon Pr[T(b) \in S]. \quad (1)$$

The definition introduces ϵ , called the privacy budget. It quantifies the level of privacy protection provided. By satisfying ϵ -Differential Privacy, a mechanism provides a strong privacy guarantee, indicating that an adversary cannot significantly differentiate between the presence or absence of an individual's data based on the mechanism's output, thereby safeguarding individual privacy during data analysis or release.

B. Randomized Response

The Randomized Response (RR) method, introduced by H. Warner et al. in 1965 [7]. With RR, when an end user is asked a binary question (e.g., "yes" or "no"), a coin is flipped with a probability of p for heads. To maintain the user's privacy, RR allows the user to provide the opposite response when heads are shown. Consequently, the data aggregator is unable to confidently ascertain the true response for a specific user, ensuring their privacy is preserved.

Definition: The RR mechanism is a mapping with $X = Y$ that satisfies the following equality:

$$Q(x|y) \begin{cases} \frac{e^\epsilon}{|Y|-1+e^\epsilon}, & \text{if } x = y \\ \frac{1}{|Y|-1+e^\epsilon}, & \text{if } x \neq y \end{cases} \quad (2)$$

Here, $Q(x|y)$ is the conditional probability, Y is the true dataset, X is the privatized dataset, $y \in Y, x \in X$, $|Y|$ is the size of set Y , and ϵ is the privacy parameter.

C. RAPPOR

Privacy-Preserving Aggregatable Randomized Response is a real world application of LDP has been made by Google for collecting statistics from the end user, and client side software, in a way that provides robust privacy protection using randomize response techniques [8]. RAPPOR's applies randomized response to bloom filters [25] with strong ϵ -differential privacy guarantees. Bloom filter is a simple space-efficient randomized data structure for representing a set in order to support membership queries.

The RAPPOR algorithm takes in the client's true value v and parameters of execution k, h, f, p, q and is executed locally on the client's machine performing the following steps:

- 1) Signal: Hash client's value v onto the bloom filter B of size k using h hash functions.
- 2) Permanent randomized response: For each client's value v and bit $i, 0 \leq i < k$ in B , create a binary reporting value B_i^v which equals to

$$B_i^v = \begin{cases} 1, & \text{with probability } \frac{1}{2}f \\ 0, & \text{with probability } \frac{1}{2}f \\ B_i, & \text{with probability } 1 - f \end{cases} \quad (3)$$

where f is a user-tunable parameter controlling the level of longitudinal privacy guarantee. Subsequently, this B_i^v is memoized and reused as the basis for all future reports on this distinct value v .

- 3) Instantaneous randomized response: Allocate a bit array S of size k and initialize to 0. Set each bit i in S with probabilities

$$P(S_i = 1) = \begin{cases} q, & \text{if } B_i^v = 1. \\ p, & \text{if } B_i^v = 0. \end{cases} \quad (4)$$

D. Machine Learning Techniques

The K-nearest neighbors (KNN) classifier is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining, and intrusion detection [26]. KNN algorithm helps us identify the nearest points or the groups for a query point. But to determine the closest groups or the nearest points for a query point we need some metric. For this purpose, we use below distance metrics:

$$d(x, y) = \left(\sum_{i=1}^n (x_i - y_i)^p \right)^{\frac{1}{p}} \quad (5)$$

Support vector machines: the support vector machines, is a powerful supervised learning algorithm used for classification tasks. It works by finding an optimal hyperplane in a high-dimensional feature space that separates different classes of data points. The hyperplane is chosen in such a way that it maximizes the margin, which is the distance between the hyperplane and the closest data points of each class [27]. This helps to achieve better generalization and robustness of the model.

Hyperplane Equation: The SVMs algorithm seeks to find a hyperplane in the feature space that separates the data points. The hyperplane equation can be written as:

$$w \cdot x + b = 0 \quad (6)$$

where:

w is a weight vector orthogonal to the hyperplane. x is the feature vector of a given data point. b is the offset or distance of the hyperplane from the origin along the normal vector w .

Classification:

$$f(x) = \text{sign}(w \cdot x + b) \quad (7)$$

where

$\text{sign}(\cdot)$ is the sign function that returns -1 or 1 depending on the sign of its argument. If $f(x) < 0$, the point is classified as one class, and if $f(x) > 0$, it is classified as the other class.

XGBoost is an implementation of gradient boosted decision trees. XGBoost models majorly dominate in many Kaggle Competitions. In this algorithm, decision trees are created in sequential form. Weights play an important role in XGBoost [28]. XGBoost objective function:

$$\text{Obj}^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)}) + \sum_{j=1}^T \Omega(f_j) \quad (8)$$

where

$\text{Obj}^{(t)}$ is the objective function at the t th iteration. n is the total number of training examples. y_i is the true label of the i th training example. $\hat{y}_i^{(t-1)}$ is the predicted value of the i th example at the $t-1$ th iteration. T is the total number of trees in the ensemble. f_j is the j th tree in the ensemble. $\Omega(f_j)$ is the regularization term that penalizes the complexity of the tree.

E. Performance Evaluation Measurements

In this paper, the classifiers performance has been analyzed by using Precision, Recall and F-measure, which are obtained from the confusion matrix as shown in Table I. These metrics are described as follows.

TABLE I. CONFUSION MATRIX

	P*(Predicted)	N*(Predicted)
P (Actual)	TP	FN
N (Actual)	FP	TN

- Precision: measures the relevant actions found against all actions found i.e. the percentage of selected actions that are correct and is defined by the following equation.

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (9)$$

- Recall: measures the relevant actions found against all relevant actions i.e. the percentage of correct actions that are selected and is defined by the following equation.

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (10)$$

- F-measure: is weighted harmonic mean between precision and recall and is defined by the following equation.

$$F\text{-measure} = \frac{2 * \text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (11)$$

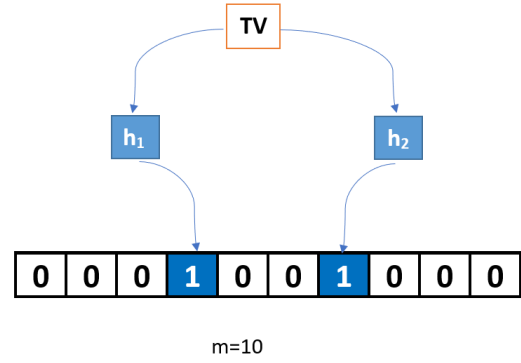


Fig. 1. Encoding algorithm maps C_i devices into bits in a bloom filter.

IV. THE PROPOSED PRIVACY-PRESERVING MODEL IN SMART HOMES

This section introduces and describe the proposed model and methodology. It outlines the key concepts and principles that underpin our approach, as well as explain the data collection process, preprocessing techniques used, and any specific algorithms or techniques used within the model.

A. Assumptions

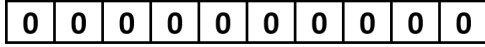
Assume that there is a smart home which contains a set of devices and each device has its own sensor to measure power consumption, and there is a set of classes $C_i, i \in \{1, \dots, l\}$ where l denotes the maximum number of supported classes, these classes represent the priority of each device, each class is composed of groups of devices D_{ij} (i.e. $j \in \{1 \dots g\}$) where g represents the maximum number of groups of interests for class i , devices like (lights, TVs, laptops, sound system, alert systems, air condition systems, laundry devices, camera systems, garden lights, garage lights/motors, fans). RAPPOR is used to send power consumption statistics to service providers/electricity products companies.

B. System Model and Overview

Based on the basic idea of bloom filter and RAPPOR, the proposed approach consists of two phases: data perturbation phase and decoding phase. These two phases are described as follows:

1) *Smart home data perturbation*: In this phase, the proposed approach determines the set of devices and each device has its own sensor to measure power consumption, and there is a set of classes $C_i, i \in \{1, \dots, l\}$ where l denotes the maximum number of supported classes, classes represent the priority of each device, each class is composed of groups of devices D_{ij} (i.e. $j \in \{1 \dots g\}$) where g represents the maximum number of devices of interests for class i . The proposed approach uses the following steps.

- 1) *Encoding* is the first step of the data perturbation process, the encoding algorithm maps C_i devices into bits in a bloom filter. Fig. 1 illustrates the bloom filter implementation using 2 hash functions, h_1 and h_2 , on the TV class.



Empty Bloom filter of size $m=10$

Fig. 2. Bloom filter B is initialized with all “0” values and size $m = 10$.

To compute the optimal bloom filter size m , given the maximum number of devices encoded into the bloom filter (*i.e.* $n = l * g$ if each class includes the same number of groups), and a fixed false positive rate f_p , according to Eq. 12:

$$m = -\frac{n \times \ln(f_p)}{(\ln 2)^2} \quad (12)$$

Then, the optimal number of hash functions is computed as given by Eq. 13

$$k = \frac{m}{n} \times \ln 2 \quad (13)$$

After selection of m and k appropriate values, the bloom filter B is initialized with all “0” values, as given in Fig. 2. For feeding B with the set of devices values v, C first applies the k hash functions to v , and feeds B with the hash output providing indices. Example: let us consider a bloom filter of size $m = 10$ bits, with $k = 2$ hash functions (h_1, h_2), and two devices {TV, Reception lights} to be included into the bloom filter. As given in Fig. 1, C first computes the two hashes of the TV device, and gets the results $h_1(TV) = 4$ and $h_2(TV) = 7$, thus leading to positioning the 4th and the 7th bits of the bloom filter to value 1. The same applies for the reception lights as depicted in Fig. 2 where h_1 (Reception lights) = 2 and h_2 (Reception lights) = 8.

- 2) *Permanent randomized response (PRR)*: This first level perturbation applies over the bloom filter B obtained through the encoding phase. This step is executed once over a set of devices v . A noisy bit is derived from each bit of B thus resulting in a perturbed bloom filter vector B' . The derivation is compliant with the RAPPOR works and considers the following probabilistic processing:

$$B'[i] = \begin{cases} 1 & \text{with probability } \frac{1}{2}f \\ 0 & \text{with probability } \frac{1}{2}f \\ B[i] & \text{with probability } 1 - f \end{cases} \quad (14)$$

- 3) *Instantaneous randomized response (IRR)*: To guarantee stronger privacy, this second level perturbation is executed for each request done by P Providers. After getting B' , the user initializes a bit vector S with all zeros and then applies the following probabilistic processing Eq. 15:

$$P(S[i] = 1) \begin{cases} q & \text{if } B'[i] = 1 \\ p & \text{if } B'[i] = 0 \end{cases} \quad (15)$$

Where p denotes the probability of flipping a bit that equals to 0 into 1 whereas q represents the probability

of keeping bits equal to 1. This second level perturbation IRR algorithm is ϵ -differential privacy with the following quantified ϵ_2 privacy budget Eq. 16:

$$\epsilon_2 = k \ln \left(\frac{q'(1-p')}{p'(1-q')} \right) \quad (16)$$

Where p' , resp. q' is the probability of observing 1 given that the same bloom filter bit was set to 0, resp. 1, as defined in the following Eq. 17 and 18.

$$p' = \frac{1}{2}fq + (1 - \frac{1}{2}f)p \quad (17)$$

$$q' = (1 - \frac{1}{2}f)(1 - q) + \frac{1}{2}f(1 - p) \quad (18)$$

- 4) Algorithm 1, shows the steps of this recognition phase:

Parameters:

- hash functions k : This is the number of hash functions used in the bloom filter. The specific value is determined by Eq. 13.
- bloom filter size m : This is the size of the bloom filter, which is determined by Eq. 12.
- privacy budget ϵ : This is a parameter related to the privacy level. Its specific value is determined by the user or the privacy configuration.

Input:

- x : This is a single row of data from a smart home, representing a specific event or measurement.
- f : This represents the privacy level configured by the homeowner. Its specific value is not mentioned in the code.

Output:

- Perturbed bloom filter vector S : This is the resulting vector after applying perturbations to the bloom filter, calculated using Eq. 15.

Steps:

- Set $x \in \mathcal{U}$ this indicates that the smart home data, represented by x , belongs to the set \mathcal{U} , which includes all available data in the smart home.
- Convert x to bloom filter vector B of size m : This step involves converting the smart home data, x , into a bloom filter vector B , of a specified size m .
- Apply permanent randomized response on B and get vector B' of size m .
- Apply instantaneous randomized response on B' and return vector S of size m .

C. Decoding Phase

In this phase, three machine learning algorithms KNN, SVMs and XGBoost were selected for their ability to work on perturbed data. Those algorithms are calibrated to fit the specification the following datasets, thus resulting into 3 configurations as detailed below:

Algorithm 1: Data perturbation

Parameter: hash functions k given by Eq. 13, bloom filter size m given by Eq. 12, privacy budget ϵ

Input: Row of smart home data x , f is the privacy level configured by home owner.

Output: Perturbed bloom filter vector S given by Eq. 15

Data: set $x \in \mathcal{U}$: \mathcal{U} is the set of all available data in smart home

```

/* Encoding */
1 Convert  $x$  to bloom filter vector  $B$  of size  $m$ 
/* PRR function */
2 Initialize an empty vector  $B'$  of size  $m$  and set all bits = 0
3 for  $i = 0$  to  $BloomFilterSize$  do
4    $B'[i] = 1$  with probability  $\frac{1}{2}f$ 
5    $B'[i] = 0$  with probability  $\frac{1}{2}f$ 
6    $B'[i] = B[i]$  with probability  $1 - f$ 
/* IRR function */
7 Initialize an empty vector  $S$  of size  $m$  and set all bits = 0
8 for  $j = 1$  to  $NumberOfhashfunctions$  do
9   for  $i = 1$  to  $BloomFilterSize$  do
10    if  $B'[i] = 1$  then
11      $S[i] = 1$  with probability  $\frac{e^{-\frac{\epsilon}{2k}}}{e^{-\frac{\epsilon}{2k}} + 1}$ 
12    else
13      $S[i] = 1$  with probability  $\frac{1}{e^{-\frac{\epsilon}{2k}} + 1}$ 
14 Return vector  $S$ ;
```

- 1) The K-nearest neighbors (KNN) classifier: is a versatile algorithm that classifies data based on the majority class of its K nearest neighbors in a training set, making it suitable for both classification and regression tasks.
- 2) Support vector machines: The Support vector machines works by finding an optimal hyperplane in a high-dimensional feature space that separates different classes of data points. The hyperplane is chosen in such a way that it maximizes the margin, which is the distance between the hyperplane and the closest data points of each class. This helps to achieve better generalization and robustness of the model.
- 3) XGBoost configuration: XGBoost is a gradient boosting algorithm. Table III gives the parameters calibrated for each dataset to optimize the model's performances. As can be shown, the configuration is slightly the same, except for parameter Subsample.

V. EXPERIMENTAL AND ANALYSIS

To evaluate the proposed approach, a real dataset The MHEALTH [29] is used. It is a data file consisting of approximately 1 million records. The data primarily consists of numerical values. Specifically, it is referred to as the "Mobile HEALTH" dataset, which captures body motion and vital signs recordings. The dataset encompasses measurements from ten

volunteers with diverse profiles while engaging in various physical activities. Also colab notebook is used. Colab [30] is a research initiative for prototyping machine learning models on powerful hardware such as GPUs and TPU. Tables II and III provide the SVMs and XGBoost parameters, as well as the KNN with $K = 3$. These machine learning algorithms are used in the proposed method to test shred data in smart home environments.

TABLE II. SVMs CONFIGURATION

Parameter	Value
SVM Type	rbf
C	1000
Gamma	0.4

TABLE III. XGBOOST CONFIGURATION

Parameter	Value
N estimators	55
Max depth	6
Min child weight	7
num rounds	10
Gamma	0.4

A. Classification Evaluation

This subsection analyses the influence of different parameters on the classification results, including the privacy budget value ϵ , the bloom filter size M and the number of hash functions k . Knowing that the accuracy for the dataset without applying LDP were KNN: 97.8%, SVMs: 98.5% and XGBoost: 98%.

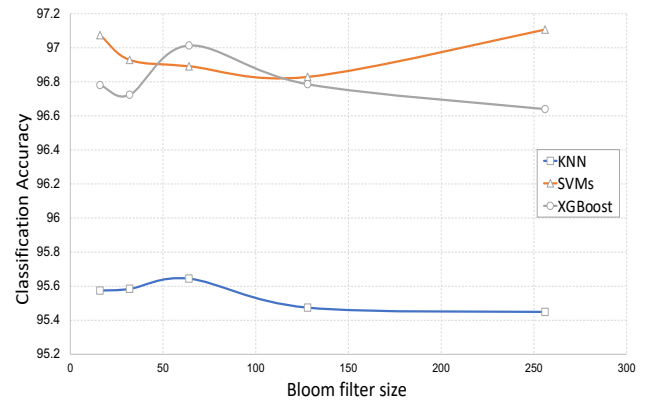


Fig. 3. Bloom filter size for $\epsilon = 0.95$ and $k = 5$.

Fig. 3 demonstrate the relationship between the accuracy of various classification methods and the size of the bloom filter. The bloom filter size ranges from 8 to 256, and it achieves accuracy within the following ranges: KNN (95.4%-95.7%), SVMs (96.8%-97.15%) and XGBoost (96.6%-97.0%). When

comparing these results with the accuracy obtained without applying LDP using the same machine learning algorithms (97.8%, 98.5%, and 98% respectively), there is an error margin of approximately 2%. However, this level of error does not significantly impact the overall accuracy.

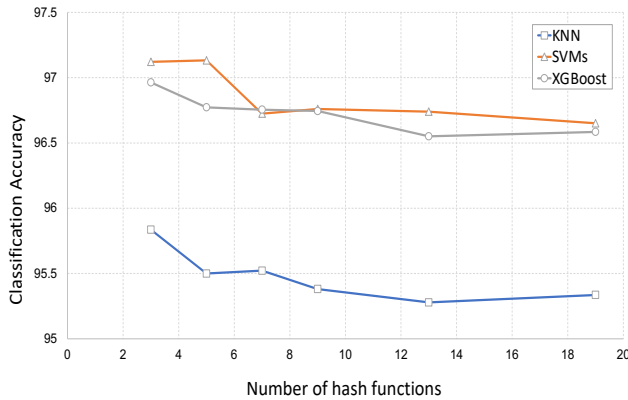


Fig. 4. Number of hash functions for $\epsilon = 0.95$ and $M = 128$].

Fig. 4 shows the relationship between the accuracy of various classification methods and the number of hash functions. The number of hash functions ranges from 3 to 19, and it achieves accuracy within the following ranges: KNN (95.2%-95.8%), SVMs (96.65%-97.13%) and XGBoost (96.55%-96.9%). Also, the error margin is approximately (1%-2%) between the proposed LDP approach and without applying LDP using the same machine learning algorithms

Our experiment considers a minimum value of hash functions of 5, which corresponds to the optimal number of hash functions for $M = 128$, according to the Eq. 13. As depicted in the Fig. 4, the classification accuracy decreases when the number of hash function increases. This stems from an increasing number of hash collisions.

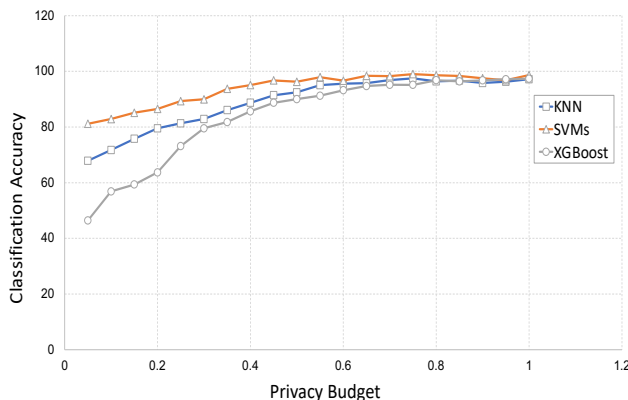


Fig. 5. Privacy budget for $M = 128$ and $K = 5$].

The relation between the accuracy of different classification algorithms and the Privacy budget is depicted in Fig. 5. The privacy budget varies from 0.1 to 1, and it achieves accuracy within the following ranges: KNN (67.2%-97.4%), SVMs

(81.65%-99.13%) and XGBoost (46.42%-97.3%). When comparing these results with the accuracy obtained without applying LDP using the same machine learning algorithms (97.8%, 98.5%, and 98% respectively).

As expected in Fig. 5, the classification accuracy is an increasing function of the privacy budget. Indeed, the higher the privacy budget, the lower the perturbation level, and the higher the accuracy. The preference dataset achieves better classification results.

B. Decoding Algorithms Evaluation

Table IV shows the accuracy of various classification methods on perturbed dataset using bloom filter size $M = 128$, privacy budget $\epsilon = 0.95$ and number of hash functions $k = 5$. Table V shows the accuracy of the same classification methods on the main dataset. As shown in both tables, an analysis of accuracy comparisons for main and perturbed data utilizing KNN, SVMs and XGBoost algorithms. This study evaluates the accuracy performance of KNN, SVMs and XGBoost algorithms when applied to a dataset consisting of 24,000 records and encompassing 12 distinct activities. The comparison focuses on the accuracy of predictions made using both the original dataset and a perturbed version.

Table VI illustrates the error margin between the accuracy of various classification methods on perturbed data and main dataset. The findings of this analysis indicate that the application of LDP techniques on the dataset did not introduce any significant impact on the decision-making process. The accuracy levels observed for the main dataset and the perturbed data remained consistent across the evaluated algorithms, namely KNN, SVMs, and XGBoost.

C. Security Analysis

In this part, we undertake security analysis using the fundamental adversary model. This model assumes that the attacker has access to the altered data disclosed by different individuals through the Local Differential Privacy (LDP) method. The primary objective is to ensure that despite the adversary's access to the perturbed data and some knowledge of the noise introduced during the LDP procedure, inferring sensitive information about any individual remains computationally infeasible or statistically improbable. The success rate of basic adversary can directly be obtained from the probability of Eq. 19 [31]

$$Pr(B'[i] = 1) \begin{cases} \frac{e^{\frac{\epsilon}{2k}}}{e^{\frac{\epsilon}{2k}} + 1} & \text{if } B[i] = 1 \\ \frac{1}{e^{\frac{\epsilon}{2k}} + 1} & \text{if } B[i] = 0 \end{cases} \quad (19)$$

Fig. 6 illustrates the relation between the success rate of basic adversary and ϵ and k values. The privacy budget ranges from 0.1 to 4, and the number of hash functions are ($k = 2, k = 7, k = 12$). This figure indicates that as the privacy budget increases, the probability that the adversary will win in the game also increases. However, as the number of hash functions increases, more wrong guesses occur. As expected, the probability of winning the game decreases when ϵ and k increase.

TABLE IV. ACCURACY FOR PERTURBED DATA USING BLOOM FILTER SIZE $M = 128$, PRIVACY BUDGET $\epsilon = 0.95$ AND NUMBER OF HASH FUNCTIONS $k = 5$, KNN, SVMs AND XGBOOST, 24,000 RECORDS AND 12 ACTIVITIES

	Precision			Recall			F1-score			Support		
	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs
Standing still (1 min)	1	1	1	1	1	1	1	1	1	1121	1121	1121
Sitting and relaxing (1 min)	1	1	1	1	1	1	1	1	1	542	542	542
Lying down (1 min)	1	1	1	1	0.99	1	1	1	1	544	544	544
Walking (1 min)	0.97	0.93	0.97	0.94	0.99	0.94	0.96	0.96	0.96	567	567	567
Climbing stairs (1 min)	0.97	0.97	0.97	0.95	0.89	0.95	0.96	0.93	0.96	610	610	610
Waist bends forward (20x)	0.99	0.99	0.99	0.98	1	0.98	0.98	0.99	0.98	567	567	567
Frontal elevation of arms (20x)	0.98	0.96	0.98	0.98	0.99	0.98	0.98	0.98	0.98	543	543	543
Knees bending (crouching) (20x)	0.97	0.95	0.97	0.98	0.96	0.98	0.98	0.96	0.98	569	569	569
Cycling (1 min)	1	1	1	1	1	1	1	1	1	581	581	581
Jogging (1 min)	0.92	0.83	0.92	0.92	0.92	0.92	0.92	0.87	0.92	576	576	576
Running (1 min)	0.96	0.93	0.96	0.94	0.88	0.94	0.95	0.9	0.95	596	596	596
Jump front & back (20x)	0.9	0.91	0.9	0.94	0.84	0.94	0.92	0.87	0.92	594	594	594
Weighted Avg	0.972	0.956	0.972	0.969	0.955	0.969	0.971	0.955	0.971	7410	7410	7410

TABLE V. ACCURACY FOR MAIN DATA USING KNN, SVMs AND XGBOOST ALGORITHMS, 24,000 RECORDS AND 12 ACTIVITIES

	Precision			Recall			F1-score			Support		
	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs
Standing still (1 min)	1	1	1	1	1	1	1	1	1	922	922	922
Sitting and relaxing (1 min)	1	1	1	0.99	0.99	0.99	1	1	1	488	488	488
Lying down (1 min)	1	1	1	1	1	1	1	1	1	450	450	450
Walking (1 min)	1	1	1	1	1	1	1	1	1	449	449	449
Climbing stairs (1 min)	1	1	1	1	0.96	1	1	0.98	1	443	443	443
Waist bends forward (20x)	1	1	1	1	1	1	1	1	1	444	444	444
Frontal elevation of arms (20x)	1	0.99	1	0.99	0.99	0.99	1	0.99	1	438	438	438
Knees bending (crouching) (20x)	1	0.96	1	1	0.99	1	1	0.97	1	432	432	432
Cycling (1 min)	1	1	1	1	1	1	1	1	1	447	447	447
Jogging (1 min)	0.98	0.92	0.98	1	0.98	1	0.99	0.95	0.99	425	425	425
Running (1 min)	1	0.98	1	0.98	0.94	0.98	0.99	0.96	0.99	458	458	458
Jump front & back (20x)	1	0.97	1	1	0.96	1	1	0.96	1	454	454	454
Weighted Avg	0.998	0.985	0.998	0.997	0.984	0.997	0.998	0.984	0.998	5850	5850	5850

TABLE VI. THE ERROR MARGIN BETWEEN THE ACCURACY OF VARIOUS CLASSIFICATION ALGORITHMS KNN, SVMs AND XGBOOST ON PERTURBED DATA AND MAIN DATASET, 24,000 RECORDS AND 12 ACTIVITIES

	Precision			Recall			F1-score		
	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs	KNN	XGBoost	SVMs
Standing still (1 min)	0	0	0	0	0	0	0	0	0
Sitting and relaxing (1 min)	0	0	0	0	0	0	0	0	0
Lying down (1 min)	0	0	0	0	0.01	0	0	0	0
Walking (1 min)	0.03	0.07	0.03	0.06	0.01	0.06	0.04	0.04	0.04
Climbing stairs (1 min)	0.03	0.03	0.03	0.05	0.07	0.05	0.04	0.05	0.04
Waist bends forward (20x)	0.01	0.01	0.01	0.02	0	0.02	0.02	0.01	0.02
Frontal elevation of arms (20x)	0.02	0.03	0.02	0.01	0	0.01	0.02	0.01	0.02
Knees bending (crouching) (20x)	0.03	0.01	0.03	0.02	0.03	0.02	0.02	0.01	0.02
Cycling (1 min)	0	0	0	0	0	0	0	0	0
Jogging (1 min)	0.06	0.09	0.06	0.08	0.06	0.08	0.07	0.08	0.07
Running (1 min)	0.04	0.05	0.04	0.04	0.06	0.04	0.04	0.06	0.04
Jump front & back (20x)	0.1	0.06	0.1	0.06	0.12	0.06	0.08	0.09	0.08
Weighted Avg	0.026666667	0.029166667	0.026666667	0.0275	0.029166667	0.0275	0.0275	0.029166667	0.0275

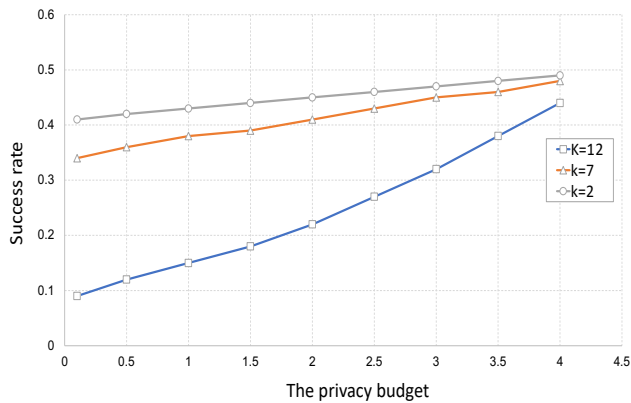


Fig. 6. Success rate over one record of perturbed data by a Basic Adversary.

VI. CONCLUSION

In this study, we investigated the problem of sharing data in a smart home environment while preserving user privacy. The main contribution of this research is the development of an efficient method for secure data sharing in smart homes using local differential privacy and the Randomized Aggregatable Privacy-Preserving Ordinal technology. Individual users' privacy is protected while data sharing with service providers is facilitated by the proposed method. The simulation results demonstrate that the technique performs well in terms of privacy preservation, accuracy, recall, and f-measure metrics, achieving utility privacy with high classification accuracy of 95.4% to 98% when the privacy budget is set to 0.95. This research helps to improving data privacy and utility in the context of smart homes, as well as providing a valuable direction for privacy-preserving practices in the IoT domain. In future research, we aim to extend the research to consider privacy-preserving techniques for multi-modal data, such as combining data from various sensors and devices within a smart home environment.

REFERENCES

- [1] T. Denning, T. Kohno, and H. M. Levy, "Computer security and the modern home," *Communications of the ACM*, vol. 56, no. 1, pp. 94–103, 2013.
- [2] H. Youssef, S. Kamel, M. Hassan, and L. Nasrat, "Optimizing energy consumption patterns of smart home using a developed elite evolutionary strategy artificial ecosystem optimization algorithm," *Energy*, p. 127793, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544223011878>
- [3] O. Taiwo, A. Ezugwu, O. Oyelade, and M. Almutairi, "Enhanced intelligent smart home control and security system based on deep learning model," *Wireless communications and mobile computing*, vol. 2022, pp. 1–22, 2022.
- [4] S. Zhang, W. Li, Y. Wu, P. Watson, and A. Zomaya, "Enabling edge intelligence for activity recognition in smart homes," in *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 2018, pp. 228–236.
- [5] G. Muhammad, M. F. Alhamid, M. Alsulaiman, and B. Gupta, "Edge computing with cloud for voice dis-

- order assessment and treatment," *IEEE Communications Magazine*, vol. 56, no. 4, pp. 60–65, 2018.
- [6] J. Lwowski, P. Kolar, P. Benavidez, P. Rad, J. J. Prevost, and M. Jamshidi, "Pedestrian detection system for smart communities using deep convolutional neural networks," in *2017 12th System of Systems Engineering Conference (SoSE)*. IEEE, 2017, pp. 1–6.
- [7] C. Dwork, "Differential privacy. automata, languages and programming-icalp 2006, lncs 4052," 2006.
- [8] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.
- [9] N. Waheed, F. Kha, M. Jan, A. Z. Alalmaie, and P. Nanda, "Privacy-enhanced living: A local differential privacy approach to secure smart home data," *arXiv preprint arXiv:2304.07676*, 2023.
- [10] R. Ratra, P. Gulia, and N. S. Gill, "Evaluation of re-identification risk using anonymization and differential privacy in healthcare," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 2, 2022. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2022.0130266>
- [11] O. Almutairi and K. Almarhabi, "Investigation of smart home security and privacy: Consumer perception in saudi arabia," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021. [Online]. Available: <http://dx.doi.org/10.14569/IJACSA.2021.0120477>
- [12] P. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827–5842, 2019.
- [13] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [14] F. McSherry and I. Mironov, "Differentially private recommender systems: Building privacy into the netflix prize contenders," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 627–636.
- [15] S. Rahali, M. Laurent, S. Masmoudi, C. Roux, and B. Mazeau, "A validated privacy-utility preserving recommendation system with local differential privacy," in *2021 IEEE 15th International Conference on Big Data Science and Engineering (BigDataSE)*. IEEE, 2021, pp. 118–127.
- [16] A. Friedman and A. Schuster, "Data mining with differential privacy," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2010, pp. 493–502.
- [17] J. Hamm, A. C. Champion, G. Chen, M. Belkin, and D. Xuan, "Crowd-ml: A privacy-preserving learning framework for a crowd of smart devices," in *2015 IEEE 35th International Conference on Distributed Computing Systems*. IEEE, 2015, pp. 11–20.
- [18] A. Mani and M. Sherr, "Histor varepsilon : Differentially private and robust statistics collection for tor." in *NDSS*, 2017.
- [19] F. Kargl, A. Friedman, and R. Boreli, "Differential privacy in intelligent transportation systems," in *Proceed-*

- ings of the sixth ACM conference on Security and privacy in wireless and mobile networks, 2013, pp. 107–112.
- [20] P. Gope and B. Sikdar, “An efficient data aggregation scheme for privacy-friendly dynamic pricing-based billing and demand-response management in smart grids,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 3126–3135, 2018.
- [21] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, “Human-factor-aware privacy-preserving aggregation in smart grid,” *IEEE Systems Journal*, vol. 8, no. 2, pp. 598–607, 2013.
- [22] H. Bao and R. Lu, “Ddpft: Secure data aggregation scheme with differential privacy and fault tolerance,” in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 7240–7245.
- [23] C. Castelluccia, A. C. Chan, E. Mykletun, and G. Tsudik, “Efficient and provably secure aggregation of encrypted data in wireless sensor networks,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 5, no. 3, pp. 1–36, 2009.
- [24] L. Lyu, K. Nandakumar, B. Rubinstein, J. Jin, J. Bedo, and M. Palaniswami, “Ppfa: Privacy preserving fog-enabled aggregation in smart grid,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3733–3744, 2018.
- [25] B. Bloom, “Space/time trade-offs in hash coding with allowable errors,” *Communications of the ACM*, vol. 13, no. 7, pp. 422–426, 1970.
- [26] V. Prasatha, H. Alfeilate, A. Hassanate, O. Lasassmehe, A. Tarawnehf, M. Alhasanattg, and H. Salmane, “Effects of distance measure choice on knn classifier performance—a review,” *arXiv preprint arXiv:1708.04321*, p. 56, 2017.
- [27] S. Yue, P. Li, and P. Hao, “Svm classification: Its contents and challenges,” *Applied Mathematics-A Journal of Chinese Universities*, vol. 18, pp. 332–342, 2003.
- [28] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [29] A. S. Oresti Banos, Rafael Garcia, “Mhealth dataset,” UCI Machine Learning Repository, 2014, dOI: <https://doi.org/10.24432/C5TW22>.
- [30] E. Bisong and E. Bisong, “Google colaboratory,” *Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, pp. 59–64, 2019.
- [31] M. E. Gursoy, A. Tamersoy, S. Truex, W. Wei, and L. Liu, “Secure and utility-aware data collection with condensed local differential privacy,” *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2365–2378, 2019.

The Implementation of Image Conceptualization Split-Screen Stitching and Positioning Technology in Film and Television Production

Zhouzhou Deng^{1*}, Rongshen Zhu²

School of Arts and Science, Chengdu College of University of Electronic Science and Technology of China,
Chengdu, 611731, China¹

School of Computer Science, Chengdu College of University of Electronic Science and Technology of China,
Chengdu, 611731, China²

Abstract—In order to study the technology of image conception, splitting, stitching and positioning in film and television production, this paper first discusses the relevant research literature, then designs an improved biomedical image segmentation convolution network model applied in film and television production, and then verifies the effectiveness of the proposed model. Ultimately, the paper summarizes the research findings. Aiming at the problem that the traditional image mosaic positioning model has poor robustness because of its insufficient ability to extract features and inaccurate segmentation and positioning areas, this study proposes a biomedical image segmentation convolutional network model that is based on dense block and void space convolutional pooling pyramidal module. Additionally, an attention mechanism is introduced to enhance the biomedical image segmentation convolutional network model. The results show that the accuracy, recall, and F1 value of the biomedical image segmentation convolutional network model are 96.48%, 95.24%, and 95.96%, respectively, on the Colombian uncompressed image stitching detection dataset, and the accuracy, recall, and F1 value of the improved biomedical image segmentation convolutional network model are 98.19%, 96.23%, and F1 value of 97.21%. In summary, the improved convolution network model for biomedical image segmentation has excellent performance, and it has certain application value in image conception, mirror splicing and positioning in film and television production.

Keywords—Convolutional neural network; attention mechanism; null space convolutional pooling pyramid; spatial rich model; dense block

I. INTRODUCTION

The field of cinema and television production has encountered novel challenges and made significant strides as a result of the ongoing advancements in science and technology [1-2]. As an essential pre-production step in film and television, the split screen is an intermediate medium for converting text into a three-dimensional audio-visual image and presenting it in a pictorial form [3-4]. Due to the increased functionality and user-friendliness of image editing software, producers can easily select areas of interest from other images to be cut and spliced into the split-screen image, which brings great convenience for creating the split-screen image, but also easily brings trouble to the film and television copyright [5]. To enhance split-screen imagery creativity in film and

television production and reduce the risk of copyright, it's crucial to prioritize research on image stitching and positioning technologies [6]. Image stitching is a semantic segmentation of features by dividing the stitching area into one category and the real area into another category. The CNBIS model is capable of extracting the stitched region from the real region. The traditional CNBIS model can achieve good results only when it is faced with simple semantic information of the same kind of content. However, in image mosaic and positioning, the mosaic area often comes from different semantic interference information, resulting in inadequate feature extraction and inaccurate segmentation and positioning area, which reduces the segmentation accuracy of the model [7]. The research aims to improve image semantic segmentation accuracy and make up for the deficiency of model extraction features caused by different semantic interference information. To overcome these issues, the study implements Dense Block (DB) and Atrous Spatial Pyramid Pooling (ASPP) modules to improve the CNBIS model and introduce the DACNBIS model. The research also incorporates the Attention Mechanism (AM) to enhance the DACNBIS model and create the AM-DACNBIS model, which increases the segmentation accuracy of the DACNBIS model. The primary contribution of the research is to broaden the application of image mosaic positioning technology in film and television production. The goal is to increase the diversity of split-mirror images, improve the reduction and richness of split-mirror in films, and make the storyline and the overall picture of film and television works more complete. The research focuses on two significant innovations. The first point is to improve the traditional CNBIS model by introducing DB module and ASPP module, and improve the DACNBIS model by combining attention mechanism. The second point is that AM-DACNBIS model is divided into three parts: parameter sharing, area monitoring and edge detection, and specific feature layers are used for each part to extract the corresponding task information. The study is primarily divided into four parts. The first part reviews relevant pertinent research findings. The second part constructs the construction of DACNBIS model and AM-DACNBIS model. The third part validates the two proposed models' validity in the study. The final part concludes the research.

II. RELATED WORK

Many image processing jobs start with the pre-processing stage of picture semantic segmentation, and many academics have written extensively about ways to increase segmentation accuracy in this process, Gao and co. To help the network better concentrate on object borders and small objects during the feature extraction process, a sensitive feature selection module was created to reweight each pixel on several channels. The findings of the experiments demonstrate that the sensitive feature selection module can aid in the semantic segmentation algorithm's high segmentation accuracy [8]. To reduce semantic information loss and improve image information, Zhou et al. proposed a semantic segmentation model based on dense convolutional separation convolution. This model also integrates multi-scale feature information for a broader perceptual field and captures more dense pixels. Simulation experiments demonstrate the superior performance and good performance of this model for image segmentation [9]. Zhang et al. proposed a system that utilizes the original semantic segmentation network to deal with the issue of missing static object segmentation occluded by dynamic scenes and combines it with image restoration [10]. Maurya et al. created a cross-form attention pyramid to extract multi-scale information using a pre-trained model in order to address the issues of feature redundancy and low discrimination in image semantic segmentation. An attention module in a spatial manner is then introduced to further improve the segmentation effect. With the addition of the attention pyramid and attention module, simulation results demonstrate that the semantic segmentation model has greater segmentation accuracy [11]. In order to tackle the problem of loss of image information caused by feature extraction in the process of image semantic segmentation, Chen constructs a semantic segmentation model with encoder-decoder as the basic structure. The research results verify the logic and effectiveness of the model [12].

CNBIS model is a very famous segmentation network model in the field of image segmentation, which has been widely used in various fields. In order to accurately divide non-enhanced tumor, enhanced tumor, tumor core and undamaged region in brain images, Teki et al. used CNBIS model to realize semantic segmentation of brain tumor images. The research findings demonstrate that the CNBIS model performs well in segmenting simple semantic information. However, in the presence of complex semantic interference information, the model exhibits insufficient feature extraction capability and low segmentation accuracy [13]. Singh et al. designed an improved Deep-CNBIS model for semantic segmentation of images using satellite images to extract vegetation cover. The simulation experiments show that the model has superior performance in semantic segmentation accuracy of satellite images [14]. Tiwari et al. proposed an improved CNBIS model for segmenting vehicles, which segmented the input images by successive encoding and decoding steps. The simulation results show that the improved CNBIS model outperforms other segmentation models in terms of segmentation accuracy [15]. Cheng et al. designed a separated convolutional CNBIS model combining

convolutional downsampling. The outcomes of the simulation experiment demonstrate that the model is capable of quickly and easily detecting fabric defects with high accuracy [16]. Mahmoud et al. developed a deep learning model based on the combination of CNBIS model. The simulation experiments demonstrate that the model performs significantly better than the traditional CNBIS model, exhibiting a higher accuracy rate [17]. Abdelraouf et al. proposed to use multi-gated expansion starting blocks. It is evident from the experimental results that the addition of multi-gate expansion start blocks improves the CNBIS model performance and demonstrates exceptional capability [18].

In summary, there are many research results on image semantic segmentation and CNBIS model applications. However, most of the image semantic segmentation studies use small data set samples, which do not sufficiently meet the deep learning requirement for extensive data training. The traditional CNBIS model has insufficient feature extraction capability, which leads to inaccurate image localization and poor robustness. The paper suggests the DACNBIS model to overcome the aforementioned issues and introduces an attention mechanism to enhance the DACNBIS model, resulting in the AM-DACNBIS model.

III. DACNBIS MODEL AND AM-DACNBIS MODEL CONSTRUCTION IN IMAGE STITCHING LOCALIZATION DETECTION

The purpose of image stitching, which separates the stitching area from the real area so that they are presented separately, can be thought of as a unique sort of semantic segmentation.

A. Construction of DACNBIS Model Based on CNBIS Model

As the initial stage of rendering text into images in film and television production, splitting is the process of dividing a film script into a series of shots that can be filmed and presented as images. Convolutional Neural Networks (CNN) are a deep learning model with learnable weights and bias constants that are suitable for processing image data processing through supervised learning [19-22]. The convolutional layer can extract features from images. Equation (1) expresses the parameters of this layer.

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} * Kernel_{ij}^l + b^l \right) \quad (1)$$

In equation (1), x_j^l is the j neuron of the l layer, $Kernel$, $*$, and $f(\square)$ represent the convolution kernel, convolution operation, and nonlinear excitation function, respectively, a and M_j represent the bias term and the number of inputs of the j neuron, respectively. The pooling layer can reduce the risk of overfitting by reducing the dimensionality. The excitation layer introduces nonlinear features to the neural network, enabling it to approximate any nonlinear function. The activation function image is shown in Fig. 1.

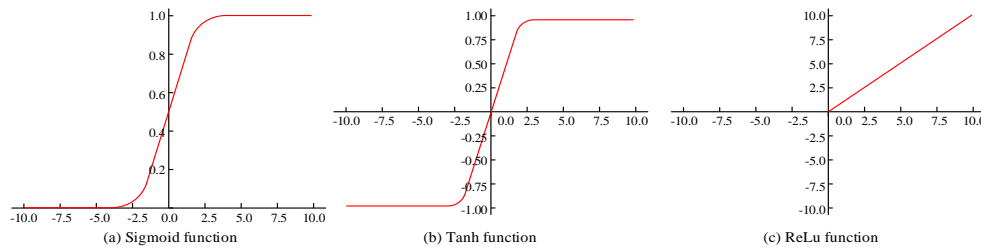


Fig. 1. Schematic diagram of three activation functions.

The Sigmoid function, Tanh function, and ReLU function, respectively, are depicted in Fig. 1 where the expression of the Sigmoid function is shown in equation (2).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

The value range of equation (2) is (0,1), which corresponds to the probability value range (0,1). Equation (3) contains the Tanh function's expression.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (3)$$

The output of equation (3) is centered on 0, which can play the effect of data centering. Due to the saturation region of Sigmoid function and Tanh function, the gradient is prone to gradient disappearance when back propagation, and this phenomenon becomes more and more obvious. This issue is mitigated by the ReLU function, which is expressed mathematically in equation (4).

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x < 0 \end{cases} \quad (4)$$

The positive activation value derivative of the ReLU function in equation (4) is 1. In order to accomplish the image classification task, the fully connected layer must integrate the input image data and map the feature map from the convolution layer into a fixed-length feature vector, as shown in equation (5).

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * w_{i,j} + b_j\right) \quad (5)$$

In equation (5), $w_{i,j}$ and b_j denote the weight and offset between neurons, respectively. CNBIS, as a special semantic segmentation algorithm in the field of deep learning, is a derivative model of CNN and capable of addressing image stitching and localization issues.

The CNBIS model in Fig. 2 adopts a fully symmetric coding-decoding structure. The coding section is downsampled four times, and the decoding link is corresponding to the four stages of the coding link, each of which incorporates the feature information corresponding to that from the coding process. This results in the final feature map is restored to the original image size. Due to the small sample of dataset about image stitching at this stage, the neural network is prone to the risk of overfitting, the study

introduces the DB and ASPP modules to improve the CNBIS model, resulting in the DACNBIS model. The DB module is a CNN with tightly connected nature. The feature map resolution of each layer is of the same size, so the channels of each layer can be stitched together in dimension. Suppose the output channel of the l layer inside the DB module is X_l , then the expression of X_l is shown in equation (6).

$$X_l = H_l[X_{l-1}, \dots, X_1, X_0] \quad (6)$$

In equation (6), H_l is the nonlinear transformation function of the l layer, $[\]$ indicates that all output feature maps of the X_0, X_1, \dots, X_{l-1} layer are combined by channel, and the expression of the network input X_i of the i layer is shown in equation (7).

$$X_i = K_0 + (i-1) \times K \quad (7)$$

In equation (7), K is the number of output channels of the nonlinear function H , and K_0 is the number of input channels. The dense jump connection implemented inside the DB module causes the number of channels after the l layer to become large. The study adopts a convolution of 1×1 before the convolution of 3×3 in the DB module. The ASPP module aims to minimize the loss of accuracy generated by the feature map in the process of recovering the resolution size of the original image. It employs four varied expansion rates of the null convolution to capture multi-scale information. The specific structure of DACNBIS model is shown in Fig. 3.

In Fig. 3, the study introduces the DB module to replace the original CNBIS model upsampled by the 3×3 convolutional network. This enhances feature reuse and propagation, increasing the richness of feature extraction. The study also introduces the ASPP module to replace the fourth sampling session of the original CNBIS model. This module extracts multi-scale information, expands the network perceptual field, and improves the model's segmentation effects on spliced regions of different sizes.

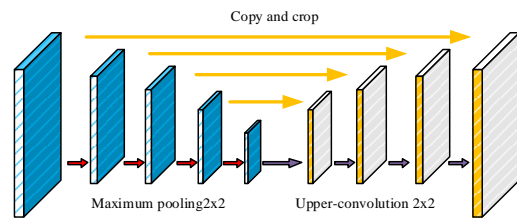


Fig. 2. Schematic diagram of CNBIS model structure.

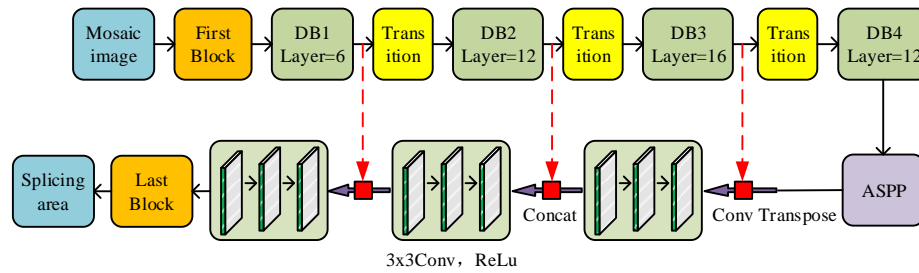


Fig. 3. Schematic diagram of DACNBIS model structure.

B. Construction of AM-DACNBIS Model based on DACNBIS Model

In order to avoid some redundant and useless low-level semantic information from interfering with the decoding process of the DACNBIS model, thereby affecting the influence of image sub-screening conception in the film and television production stage, the research combines AM to improve the DACNBIS model, resulting in the AM-DACNBIS model. It is based on the principle that humans selectively focus on the more interesting and informative visual areas when observing a scene, thus ignoring other irrelevant areas and thus improving the utilization of visual information. The study draws inspiration from AM and introduces a Global Attention Upsampling (GAU) module to provide global context as a low-level guide, see Fig. 4 [20].

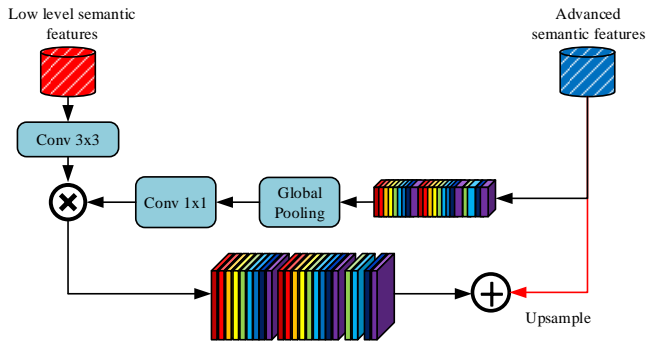


Fig. 4. GAU module schematic diagram.

In Fig. 4, GAU first convolves the low-level semantic features with 3×3 . Since the AM-DACNBIS model ignores the differences between the essential attributes of images when performing image input, which in turn leads to large differences between the extracted real images and the stitched images and affects the splitting effect in the film and TV production process. In order to reduce the interference situation of different semantic information in the stitching region on the network's extracted features, the study implements the Spatial Rich Model (SRM) filter in the input stage to analyze the image features. SRM is a high-latitude steganographic model, when the image hides secret information, SRM will destroy the image attributes and extract the feature information from these images that are hidden information. Suppose the image pixel is X , the expression of image residual $R_{i,j}$ is shown in equation (8).

$$j = 1, 2, \dots, n_2, i = 1, 2, \dots, n_1, R_{i,j} = \hat{X}_{i,j}(N_{i,j}) - cX_{i,j} \quad (8)$$

In equation (8) n_1, n_2 denote the pixel points in horizontal and vertical directions respectively, $N_{i,j}$ is the field without the central pixel $X_{i,j}$, $\hat{X}_{i,j}(N_{i,j})$ refers to the estimated value of the central pixel $X_{i,j}$ in the region $N_{i,j}$, c is the residual order, when the magnitude of the residual pixel is large, the pixel correlation there will be reduced, so the study also needs to quantize and truncate the residual image, as shown in equation (9).

$$R_{i,j} \leftarrow Trunc_T \left(\text{round} \left(\frac{R_{i,j}}{Q} \right) \right), j = 1, 2, \dots, n_2, i = 1, 2, \dots, n_1 \quad (9)$$

In equation (10), Q is the quantization step, round indicates rounding, and $Trunc_T$ means the elements are truncated one by one by the threshold T , where $Trunc_T(x)$ is expressed as shown in equation (10).

$$Trunc_T(x) = \begin{cases} x, & x \in [-T, T] \\ T \text{sign}(x), & x \notin [-T, T] \end{cases} \quad (10)$$

In equation (10) $\text{sign}(x)$ is the symbolic function and the high-pass filter extracted from the SRM can be used for the extraction of RGB image noise, as shown in equation (11).

$$K = \frac{1}{2} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (11)$$

The residual image resulting from the extraction of the SRM filter in equation (11) highlights edge features of the stitched region while suppressing other contents. In order to maintain sharing in the feature extraction process and minimize the risk of network overfitting, the study comprises of a multi-task learning output from the branching task of stitching edge localization using the DACNBIS model. According to Fig. 5, the hard parameter sharing and soft parameter sharing categories best describe the multitask learning structure.

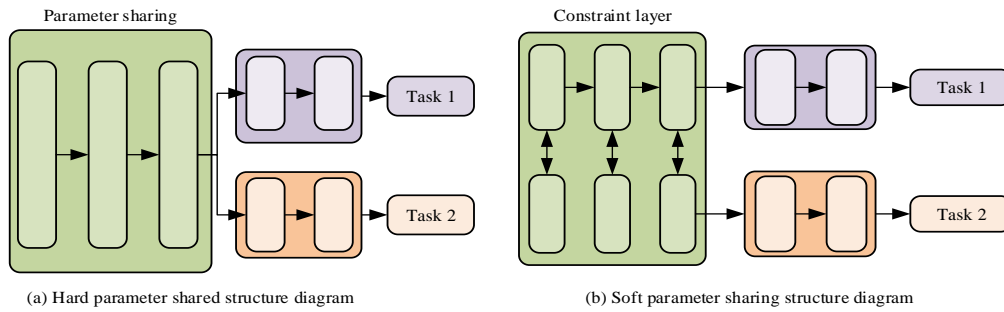


Fig. 5. Schematic diagram of multi-task learning structure.

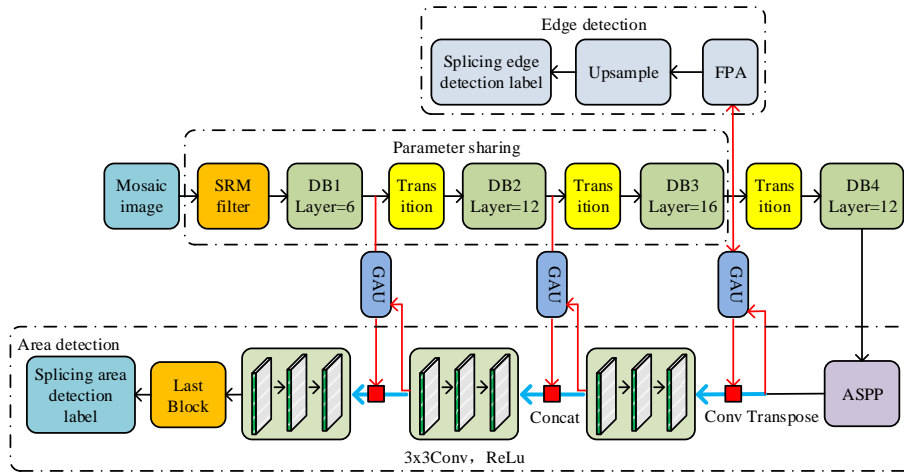


Fig. 6. Structural diagram of AM-DACNBIS model.

Fig. 5(a) shows the hard parameter sharing structure, in which the bottom layers of the network's input share parameters. This is known as bottom parameter sharing. Different learning tasks present different branches after sharing, and these tasks are trained in parallel with each other, and the feedback action is performed through the loss function each learning task. Fig. 5(b) shows the soft parameter sharing structure, in which every task has independent models and parameters. Every model can access the internal information of other models and regularize the distance between model parameters to ensure the similarity between parameters. Since the soft parameter sharing mechanism has separate models among multiple tasks. A schematic representation of the final AM-DACNBIS model structure is shown in Fig. 6. The study also embeds the data information of multiple tasks into a single semantic space and extracts the relevant task information for each task through a specific feature layer.

The model structure in Fig. 6 is mainly divided into three parts: parameter sharing, region monitoring, and edge detection. For the parameter sharing part, the AM-DACNBIS model adds SRM filters and completes the shared training for both region detection and edge detection tasks through the DB module in the pre-input phase of the network. It then embeds the information of both tasks into the same semantic space. For the region detection part, the study introduces the GAU module in the decoding process of the AM-DACNBIS model to provide global context as the underlying guidance to improve the sensitivity of important feature information. For the edge detection part, the study invokes the feature pyramid

attention module in the edge branch and learns better feature representation at the parameter sharing layer through this module. To evaluate the AM-DACNBIS performance, the study presents the performance metrics of accuracy, recall, and F1 value for testing, as shown in equation (12).

$$\begin{cases} R = \frac{TP}{FN + TP} \\ F1 = \frac{2PR}{R + P} \\ P = \frac{TP}{FP + TP} \end{cases} \quad (12)$$

In equation (12), P and R represent the precision and recall, respectively. TP is the actual stitched area pixels and the predicted stitched area pixels; FP is the actual real area pixels and the predicted stitched area pixels; FN is the actual stitched area pixels and the predicted real area pixels.

IV. ANALYSIS OF THE RESULTS OF DACNBIS MODEL AND AM-DACNBIS MODEL IN IMAGE STITCHING LOCALIZATION DETECTION

The section focuses on the examination of the experimental results of the DACNBIS model and the AM-DACNBIS model. To confirm the validity of the DACNBIS model and the AM-DACNBIS model, comparative tests were conducted on various data sets utilizing different model sets.

A. Experimental Data Preparation

The algorithmic model was developed with the PyTorch deep learning framework, and the hardware environment for the experiments was a workstation running Windows OS. This was done to compare the performance of the proposed models. With a stochastic gradient descent network training optimizer, a binary cross-entropy loss function, an initial learning rate of 0.01, momentum of 0.9, weight decay of 0.0005, and performance metrics of accuracy, recall, and F1 value. The Chinese Academy of Sciences Institute of Automation 1 (CASIA1) dataset, CASIA2 dataset, Columbia Uncompressed Image Splicing Detection (CUIS) dataset, and the Chinese Academy of Sciences Institute of Automation 1 (CASIA2) dataset are used for the study. For the experiments, the Image Splicing Detection (CUISD) datasets are utilized, and the specific experimental data are divided as shown in Table I.

TABLE I. PARTITION RESULT OF DATA SET

Data Set	CASIA1	CASIA2	CUISD
Training set	1050	7000	175
Verification set	300	2000	50
Test set	150	1000	25

1500 samples total for the CASIA1 dataset, 10,000 samples total for the CASIA2 dataset, and 250 samples total for the CUISD dataset are shown in Table I, where the training, validation, and test sets account for 70%, 20%, and 10% of the corresponding datasets. To enhance the understanding of the datasets, the schematic diagram of some data sets is selected, as shown in Fig. 7.

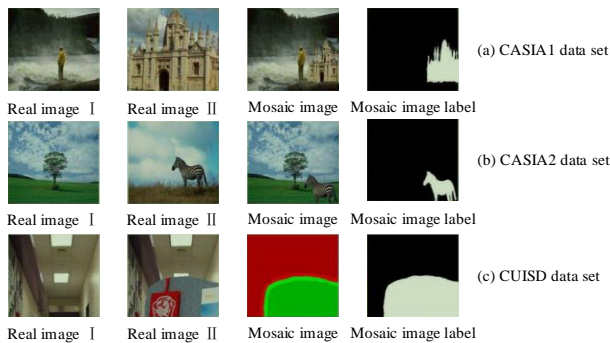


Fig. 7. Sample diagrams of three data sets.

Fig. 7(a) is a schematic diagram of the CASIA1 dataset with stitching areas of different sizes and arbitrary boundaries such as circles, triangles and rectangles. Additionally, Fig. 7(b) illustrates the CASIA2 dataset, which is an upgrade of the CASIA1 dataset with more data and better production. Fig. 7(c) displays the CUISD dataset which provides labels with red and green colored edge templates, and the edge labeling error is larger.

B. Performance Analysis of DACNBIS Model

To verify the performance of DACNBIS model, the study conducted comparison experiments using Fully Convolutional

Networks (FCN), CNBIS, and Pyramid Scene Analysis Network (PSAN) models. The number of model iterations was increased from 10 to 100 in epoch.

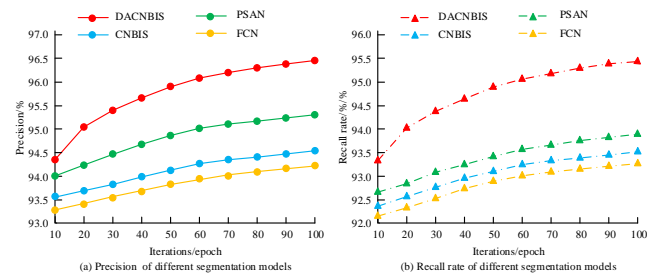


Fig. 8. Accuracy and recall results of four models.

The change curves indicated in Fig. 8 illustrate the precision rate and recall rate of the four models. As the number of iterations increases, the precision rate and recall rate of the four models improve. In Fig. 8(a), the accuracy rate for the four models is displayed, showing that the DACNBIS model has a precision rate of 96.48%. The precision rates for the PSAN, CNBIS, and FCN models are 94.87%, 94.23%, and 93.75%, respectively, at 100 epochs. Fig. 8(b) shows the recall variation curves of the four models, and the recall rates of DACNBIS, PSAN, CNBIS, and FCN models are 95.24%, 93.52%, 93.17%, and 92.91%, respectively, when the number of iterations is 100 epoch.

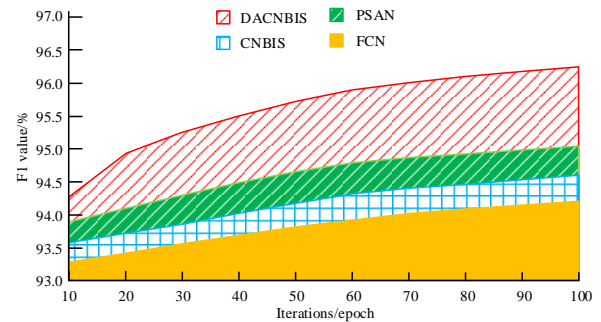


Fig. 9. F1 value results of four models.

Fig. 9 shows the results of the F1 values of the four models, which all increase with the number of iterations. When the number of iterations is 10 epoch, the F1 values of the DACNBIS, PSAN, CNBIS, and FCN models are 94.30%, 93.81%, 93.54%, and 93.26%, respectively. When the number of iterations is 100epoch, the F1 value of DACNBIS model is 95.96%, surpassing the 94.58% of the PSAN model, the 93.99% of the CNBIS model, and the 93.71% of the FCN model. In summary, the DACNBIS model proposed in the study has higher segmentation accuracy compared to other models and performs well in the field of split-screen image stitching localization in film and TV production.

C. Performance Analysis of AM-DACNBIS Model

In order to verify the validity of AM-DACNBIS model, literature [14], DACNBIS model and AM-DACNBIS model were set up for comparative experiments.

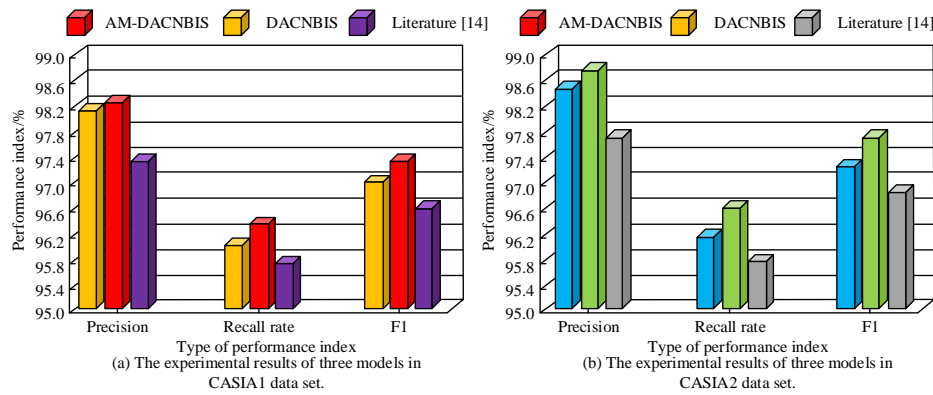


Fig. 10. The experimental results of three models in CASIA1 and CASIA2 data sets.

Fig. 10 shows the experimental results of three models in CASIA1 and CASIA2 data sets. Fig. 10(a) shows the results of accuracy, recall and F1 value of the three models on the CASIA1 data set. The accuracy of AM-DACNBIS model is 98.19%, and that of study [14] and DACNBIS model is 97.21% and 97.97% respectively. The recall rates of AM-DACNBIS, study [14] and DACNBIS models are 96.23%, 95.68% and 95.87% respectively, and the F1 value of the AM-DACNBIS model is 97.21%, which is higher than that of study [14] and DACNBIS model, which is 96.57% and 96.92%. Fig. 10(b) shows the results of accuracy, recall and F1 value of the three models in CASIA2 data set. The accuracy of AM-DACNBIS model is 98.76%, and that of study [14] and the DACNBIS model is 97.59% and 98.45% respectively. The recall rates for the AM-DACNBIS, study [14] and DACNBIS models are 96.54%, 95.51% and 96.01% respectively. The F1 value for the AM-DACNBIS model is 97.65%, which is higher than that of study [14] and the DACNBIS model, which are 96.73% and 97.23%.

Fig. 11 shows the detection results of the DACNBIS model and AM-DACNBIS model on two data sets; Fig. 11(a) displays the outcomes for both models in the CASIA1 dataset. The detection results for both models in the CASIA2 dataset are shown in Fig. 11(b). The figure shows that the AM-DACNBIS model has better performance in detection and can significantly amplify the sensitivity of key features, and can filter out key feature information. To further verify the robustness of the AM-DACNBIS model, the study performs two operations of compression and Gaussian blurring on the CASIA1 test set images, where the image compression factors

are 95, 90, 80, and 70, and the standard deviations of Gaussian blurring are set to 0.5, 1.0, 1.5, and 2.0. The results of the model tests are shown in Table II.

Table II shows that the accuracy, recall, and F1 values of both models decrease as the picture compression factor drops and the Gaussian fuzzy standard deviation rises. The AM-DACNBIS model's accuracy, recall, and F1 values are 76.58%, 59.63%, and 62.99%, respectively, when the image compression factor is 70, while those of the DACNBIS model are, respectively, 70.28%, 40.21%, and 50.31%. When the Gaussian fuzzy standard deviation is 2.0, the accuracy rate of AM-DACNBIS and DACNBIS models are 85.07% and 78.74%, the recall rates of AM-DACNBIS and DACNBIS models are 85.41% and 74.31%, and the F1 values of AM-DACNBIS and DACNBIS models are 85.21% and 76.64%. The combined results show that AM-DACNBIS can effectively reduce the risk of overfitting and is more suitable for split-screen image conception in the film and television production than the DACNBIS model.

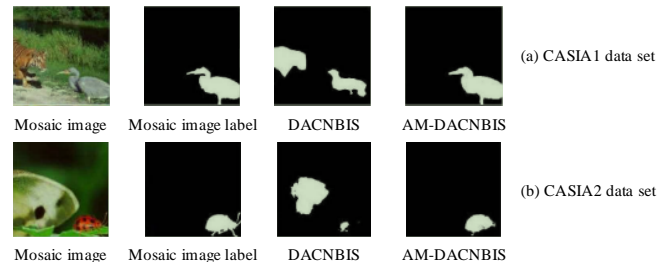


Fig. 11. Test results of two models in CASIA1 and CASIA2 data sets.

TABLE II. DETECTION OF TWO MODELS UNDER DIFFERENT COMPRESSION FACTORS AND GAUSSIAN FUZZY STANDARD DEVIATION ATTACKS

Model Type		AM-DACNBIS			DACNBIS		
Performance index		Precision	Recall	F1	Precision	Recall	F1
Compressibility factor	95	87.95%	78.27%	82.81%	87.03%	70.54%	77.74%
	90	86.94%	67.61%	75.94%	80.51%	62.41%	70.24%
	80	81.47%	60.41%	64.12%	79.74%	44.09%	55.89%
	70	76.58%	59.63%	62.99%	70.28%	40.21%	50.31%
Gaussian fuzzy standard deviation	0.5	87.41%	85.54%	86.48%	86.31%	78.79%	82.45%
	1.0	87.02%	86.21%	85.84%	84.29%	77.59%	80.81%
	1.5	86.83%	85.77%	85.64%	80.87%	75.47%	78.54%
	2.0	85.07%	85.41%	85.21%	78.74%	74.31%	76.64%

V. CONCLUSION

As image editing software becomes more functional and easier to use, the stitching and positioning techniques for splitting images, a crucial step in pre-production for film and TV, is imperative. The traditional stitching and positioning techniques have limited segmentation accuracy and poor robustness due to inadequate feature extraction capabilities in the sampling process and inaccurate positioning segmentation for different shapes. To address the above problems, a DACNBIS model based on DB module and ASPP module is proposed. The results show that when the number of iterations is 100 epoch, the accuracy, recall and F1 values of DACNBIS model under CUISD dataset are 96.48%, 95.24% and 95.96%, respectively, which are higher than 94.87%, 93.52% and 94.58% of PSAN model. Under the CASIA2 dataset, the AM-DACNBIS model exhibited higher accuracy, recall, and F1 values than the DACNBIS model with scores of 98.76%, 96.54%, and 97.65%, respectively, which were improved by 0.31%, 0.53%, and 0.42%. Under the CASIA2 dataset, the F1 values of the AM-DACNBIS model are 62.99% and 85.21% when the image compression factor and Gaussian fuzzy standard deviation are 70 and 2.0, respectively, which are higher than those of the DACNBIS model by 50.31% and 76.64%. In summary, the DACNBIS model proposed in the study performs well with the AM-DACNBIS model, but in the CASIA1 and CASIA2 datasets, the AM-DACNBIS model performs significantly better than the DACNBIS model and is more suitable for the image splitting conceptualization applications in film and television production. However, there are still shortcomings in the study, and the image quality of the CUISD, CASIA1, and CASIA2 datasets is somewhat different from the demand of multi-scope image stitching, and the subsequent study will further construct a high-quality and complex professional stitching dataset.

REFERENCES

- [1] Yang Y, Song X. Research on face intelligent perception technology integrating deep learning under different illumination intensities. *Journal of Computational and Cognitive Engineering*, 2022, 1(1):32-36.
- [2] Kiani F, Nematzadeh S, Anka F A, Findikli M A. Chaotic Sand Cat Swarm Optimization. *Mathematics*, 2023, 11(10): 2340.
- [3] Li J, Wang J. Digital animation multimedia information synthesis based on mixed reality framework with specialized analysis on speech data. *International Journal of Speech Technology*, 2023, 26(1):63-76.
- [4] Tyagi S, Yadav D. A detailed analysis of image and video forgery detection techniques. *The Visual Computer*, 2023, 39(3):813-833.
- [5] Olson E O. SCANNER IMAGING OF COMMON SHARP-TAILED SNAKES (CONTIA TENUIS) FOR INDIVIDUAL IDENTIFICATION. *Northwestern Naturalist*, 2023, 104(1):26-36.
- [6] Tang Z, Ma G, Lu J, Wang Z, Fu B, Wang Y. Sonar image mosaic based on a new feature matching method. *IET Image Processing*, 2020, 14(10):2149-2155.
- [7] Su Z, Li W, Ma Z, Gao R. An improved U-Net method for the semantic segmentation of remote sensing images. *Applied Intelligence*, 2022, 52(3): 3276-3288.
- [8] Gao Y, Che X, Liu Q, Bie M, Xu H. SFSM: sensitive feature selection module for image semantic segmentation. *Multimedia Tools and Applications*, 2023, 82(9):13905-13927.
- [9] Zhou E, Xu X, Xu B, Wu H. An enhancement model based on dense atrous and inception convolution for image semantic segmentation. *Applied Intelligence*, 2023, 53(5):5519-5531.
- [10] Zhang J, Liu Y, Guo C, Zhan J. Optimized segmentation with image inpainting for semantic mapping in dynamic scenes. *Applied Intelligence*, 2023, 53(2):2173-2188.
- [11] Maurya A, Chand S. Cross-form efficient attention pyramidal network for semantic image segmentation. *AI Communications*, 2022, 35(3):225-242.
- [12] Chen Y. Semantic Image Segmentation with Feature Fusion Based on Laplacian Pyramid. *Neural Processing Letters*, 2022, 54(5):4153-4170.
- [13] Teki S M, Varma M K, Yadav A K. Brain Tumour Segmentation Using U-net Based Adversarial Networks. *Traitement du Signal*, 2019, 36(4):353-359.
- [14] Singh N J, Nongmeikapam K. Semantic segmentation of satellite images using deep-unet. *Arabian Journal for Science and Engineering*, 2023, 48(2):1193-1205.
- [15] Tiwari T, Saraswat M. A new modified-unet deep learning model for semantic segmentation. *Multimedia Tools and Applications*, 2023, 82(3):3605-3625.
- [16] Cheng L, Yi J, Chen A, Zhang Y. Fabric defect detection based on separate convolutional UNet. *Multimedia Tools and Applications*, 2023, 82(2):3101-3122.
- [17] Mahmoud A S, Mohamed S A, El-Khoriby R A, AbdelSalam H M, El-Khodary I A. Oil Spill Identification based on Dual Attention UNet Model Using Synthetic Aperture Radar Images. *Journal of the Indian Society of Remote Sensing*, 2023, 51(1):121-133.
- [18] Abdelraouf D, Essam M, Elattar M. Light-Weight Localization and Scale-Independent Multi-gate UNET Segmentation of Left and Right Ventricles in MRI Images. *Cardiovascular Engineering and Technology*, 2022, 13(3):393-406.
- [19] Zhang Y, Lambert M, Fraysse A, Lesselier D. Unrolled convolutional neural network for full-wave inverse scattering. *IEEE Transactions on Antennas and Propagation*, 2022, 71(1):947-956.
- [20] Cong S, Zhou Y. A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review*, 2023, 56(3):1905-1969.
- [21] Nematzadeh S, Kiani F, Torkamanian-Afshar M, Aydin N. Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases. *Computational biology and chemistry*, 2022, 97: 107619.
- [22] Peng D, Yu X, Peng W, Lu J. DGFAU-Net: Global feature attention upsampling network for medical image segmentation. *Neural Computing and Applications*, 2021, 33(18):12023-12037.

Rural Landscape Design Data Analysis Based on Multi-Media, Multi-Dimensional Information Based on a Decision Tree Learning Algorithm

Ning Leng¹, Hongxin Wang^{2*}

Guangdong Peizheng College, Guangdong, China¹
University of Sanya, Hainan, China^{2*}

Abstract—This paper analyzes and studies the design characteristics of multi-dimensional information rural scenes. For data mining and the Decision Tree (DT) calculation method, the pre-processing system and method of multi-dimensional information rural award design are put forward again. Through the analysis of the multi-dimensional value of multi-dimensional multimedia mountain villages, the form of planning and design analysis and corresponding methods are based on the analysis. Using one village as a case study, we were able to investigate the villagers, roads, services, greening, ecology, and other aspects of the village in complete detail and then implement the multi-media, multi-resource village's detailed planning and design.

Keywords—Multi-media multi-dimensional information rural landscape; data mining; decision tree; data preprocessing

I. INTRODUCTION

In recent years, with the rapid development of building automatic control systems and building management systems, which have been widely used (such as large-scale public multimedia multidimensional information rural scenery supervision and management platform and landscape measurement project successively conducted by several provinces and cities in China), extensive multimedia multidimensional information rural scenery design resources are loaded. These resources are the most direct data on the operation of buildings, systems, and equipment and are highly valued. Cities nationwide have been paying attention to the call to reform rural areas in recent years [1-3]. A loss of the original rural characteristics develops in some cities because of rural urbanization, which impacts the landscape by destroying old buildings and building high-rise structures. A decline in rural scenes has resulted from cultural pressures, resulting in a loss of rural characteristics. Destruction of ecosystems directly results from financial growth and poor resource management in rural areas [4-7]. *E.g.*, reducing land in rural fields and reducing natural spot areas. Rural ecosystems are damaged by the excessive use of chemicals such as pesticides, herbicides, and fertilizers, and the township businesses' rapid processing also increases pollution.

¹This research study is sponsored by these projects: project one: The Young Innovative Talents Project (Humanities and Social Sciences) of Guangdong Province, the project number is: 2019WQNCX127. Project two: Philosophy and Social Sciences "the 13th Five-Year Plan" Youth Project of Guangdong Province, the project number is: GD20YYS05. Project three: 2020-2021 academic year scientific research project of Guangdong Peizheng College; the project number is: pzxjzd02.Thank these projects for supporting this article!

In order to obtain high-quality mining data, the results of the data to guide practical operation through the construction of a data pretreatment system, using the Decision Tree (DT) method to calculate the missing data of construction data, abnormal data, multi-dimensional data for data analysis, data pretreatment system and methods in the practical case, to obtain superior results.

II. FEATURES OF BIG DATA AND DATA PREPROCESSING SYSTEM OF MULTI-MEDIA MULTI-DIMENSIONAL INFORMATION RURAL LANDSCAPE (MMMDIRL) DESIGN

It is already under study; the relevant research on the application of data mining in architecture focuses on regular mining using data mining algorithms. MMMDIRL analysis, landscape and load prediction, fault diagnosis, optimization of operation control, and other data pretreatment research and results are limited. This section will refer to the data preprocessing processes and methods in other fields (*e.g.*, big Internet data.) and propose the preprocessing system and methods for building data after analyzing the characteristics of multi-dimensional information rural landscape data in multimedia.

DT learning algorithm variables are represented by a triple array (r_1, r_2, r_3) $r_1 < r_2 < r_3$ consisting of zeros, whose dependent function is EQU (1).

$$\mu(x) = \begin{cases} \frac{x - r_1}{r_2 - r_1}, & \text{if } r_1 \leq x \leq r_2 \\ \frac{x - r_3}{r_2 - r_3}, & \text{if } r_2 \leq x \leq r_3 \\ 0, & \text{others} \end{cases} \quad (1)$$

Set DT learning algorithm $\alpha = (a_1, a_2, a_3)$, $\beta = (b_1, b_2, b_3)$ according to the DT learning algorithm number of addition and extension principle of power, EQU (2).

$$\mu_{\alpha+\beta}^-(z) = \sup\left\{\min\left\{\mu_{\alpha}^-(x), \mu_{\beta}^-(y)\right\} \mid z = x + y\right\} = \begin{cases} \frac{z - (a_1 + b_1)}{(a_2 + b_2) - (a_1 + b_1)}, & a_1 + b_1 \leq z \leq a_2 + b_2 \\ \frac{z - (a_3 + b_3)}{(a_2 + b_2) - (a_3 + b_3)}, & a_2 + b_2 \leq z \leq a_3 + b_3 \\ 0, & \text{others} \end{cases} \quad (2)$$

That is, the sum of the DT learning algorithm is the DT learning algorithm, and there are EQ. (3) and EQ. (4)

$$\bar{\alpha} + \bar{\beta} = (a_1 + b_1, a_2 + b_2, a_3 + b_3) \quad (3)$$

From $\mu_{\lambda\bar{\alpha}}^-(z) = \sup\{\mu_{\alpha}^-(x) \mid z = \lambda x\}$
get

$$\lambda\bar{\alpha} = \begin{cases} (\lambda a_1, \lambda a_2, \lambda a_3), & \lambda \geq 0 \\ (\lambda a_4, \lambda a_3, \lambda a_2), & \lambda < 0 \end{cases} \quad (4)$$

It $\bar{\alpha}_i = (a_{i1}, a_{i2}, a_{i3}) \quad i = 1, 2, \dots, m$ is the DT learning algorithm, the non-negative linear combination and DT learning algorithm programming $\bar{\alpha}_i$ are obtained, EQ. (5)

$$\sum_{i=1}^m \lambda_i \bar{\alpha}_i, \lambda_i \geq 0 \quad (5)$$

It is still a DT learning algorithm, and EQ. (6)

$$\sum_{i=1}^m \lambda_i \bar{\alpha}_i = \left(\sum_{i=1}^m \lambda_i a_{i1}, \sum_{i=1}^m \lambda_i a_{i3} \right) \quad (6)$$

The DT learning algorithm is a random plan with constraints, such as random parameters, and chance is used to show precisely how probable it is that constraints will be set. A constraint programming environment enables the use of possibilities as constraints [8-10]. Random DT learning algorithm and DT learning algorithm control plan are powerful tools for solving optimization problems with random parameters and DT learning algorithm parameters. Compared with other industries, due to the early construction of landscape supervision and management platforms, design points are designed and evaluated in advance in the process of scheme customization. At the same time as data redundancy, the pre-set data attributes and accuracy in the database of the landscape management platform are avoided as far as possible. Problems such as incomplete data accuracy and inconsistent format will occur, so data redundancy, format contradiction, and accuracy are problematic [11-13].

Based on the above analysis, this paper proposes a pre-processing system of MMMDIRL design big data based on the characteristics and corresponding processing methods of MMMDIRL design big data, as shown in Fig. 1.

A. Significance of Rural Landscape Planning (RLP) and Design

The purpose of RLP and design is to optimize the environment of human settlements. At the same time, it is the direction of the future development of the rural environment. It is the ideal structure of the rural ecosystem. The primary purpose of RLP is to coordinate rural landscape and ecological development.

To make the relationship between man and nature more harmonious, China's new socialist rural construction is still in its infancy. RLP is the exploration and attempt of rural construction mode and the key to determining the quality of rural residents' environmental construction.

The key point is that RLP aims to optimize the rural environment and meet local ecological needs. The functions of rural landscape construction are as follows:

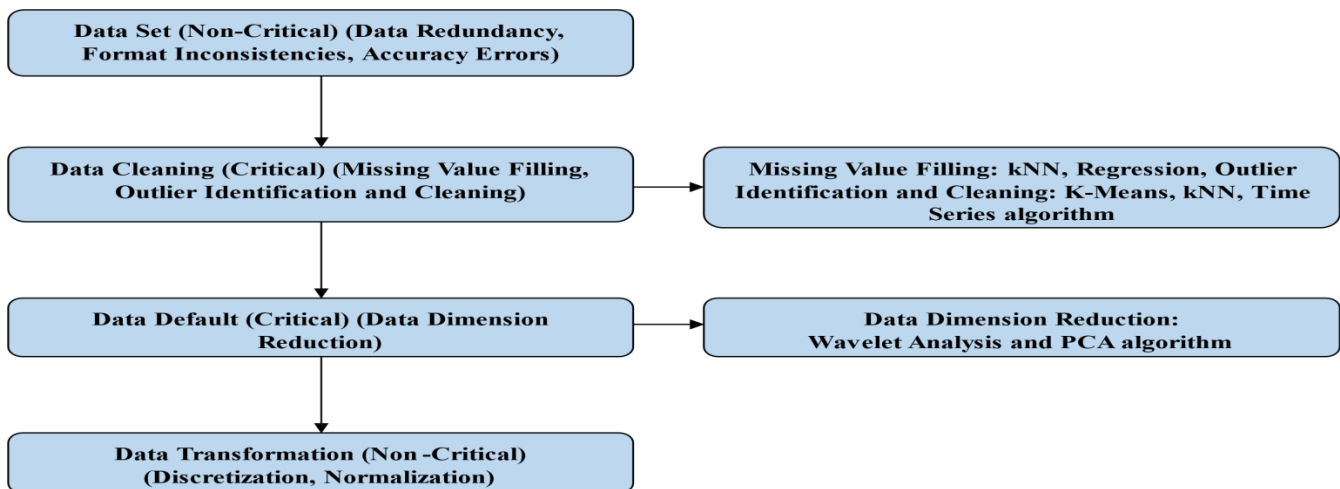


Fig. 1. Big data preprocessing technology system of MMMDIRL design.

B. Promoting Rural and Local Economic Development

RLP and design optimize and adjust the rural industrial structure, promoting the development of the local economy. The practice results in many rural areas show that while optimizing the industrial structure, the regional economy is multiplied, and the output of various crops will increase. Taking Lijiahe Township, Xuanen County, Hubei Province as an example, the adjustment of industrial structure has significantly increased the income of tobacco farmers and brought considerable tax revenue to the local finance. Another example is the "Farmhouse Fun" Ecological Sightseeing Park, which mainly displays local folk customs and promotes agricultural development by providing leisure and entertainment activities, which has extensively promoted the construction of local rural areas.

RLP focuses on the overall planning of the rural environment to realize the transformation from design to reality. RLP and design consider the balance between man and land, economic development, and environmental protection.

C. Conducive to the Inheritance and Development of National Culture

RLP is conducive to the layout of rural villages and the protection of local national culture. For example, local architectural concepts and literary works can inherit national cultural characteristics. The local architectural landscape is the carrier of national culture and fully reflects the local folk characteristics. Therefore, the rural landscape is the most essential way for people to get familiar with the traditional culture, protect the rural characteristics to the greatest extent and realize the inheritance of regional culture.

III. BUILDING DATA PREPROCESSING TECHNOLOGY BASED ON THE DT LEARNING ALGORITHM

Because of the characteristics of multimedia multi-dimensional information rural landscape design big data, in the above mentioned, the identification of abnormal data is K-ManASKNN, such as time series and other algorithms. This paper uses only one specific method to pre-process the related data quality problems; fill in missing data, determine the DT learning algorithm, and clean outliers to reduce data volume, identify outliers, and determine the DT learning algorithm.

IV. DATA QUALITY ANALYSIS

Infrastructure built on multimedia landscape and electronic network maps is well-known and durable. 200 multimedia design landscapes that won professional recognition were used as a statistical sample. The formation mechanism of the landscape's ornamental nature was assumed as the classification benchmark. The sample of 100 landscape cases has 17 active multi-media landscapes; 17% of the total. While ornamental landscapes have 78 active and passive loads, passive loading in multimedia landscape architecture causes a subject of 6% of the overall. The prototype spatial interface pattern is controlled in multiple dimensions, primary visual data is gathered at eye level, and stable triangular composition is maintained to achieve 78% of the total. The absolute

proportion of multimedia landscape architecture with a multi-dimensional dominant space interface is 85%. Data quality is assessed before mining in MMDIRL designs, which frequently use plain ornamental subjects (Table I).

TABLE I. SHOWS THE ACTUAL RUNNING DATA OF THE COUNTRYSIDE WITH MULTI-MEDIA, MULTI-DIMENSIONAL INFORMATION

S/N	P/kW	Q_{ch} /kW	T_{ci} /°C	T_{co} /°C	T_{ei} /°C	T_{eo} /°C
1	35.5	54.9	20.4	20.7	14.3	12.7
2	35.5	54.9	21.3	21.6	13.3	11.8
3	37.6	54.9	23.1	23.1	13.5	12
...
35378	97.8	504.7	28.5	33	16	10.2
35379	97.9	536.8	27.3	32.4	18.7	11.5
35380	99.5	536.8	27	31.8	21.8	14.8

A. Missing Data

The data from 35,380 groups are classified and statistically analyzed according to their attributes. Data of different attributes are missing to their extent. The degree of missing data attributes is shown in Table II. According to the analysis of data in the Table II, the missing data of the mechanical landscape attribute is the largest, with 576 missing data, accounting for 1.6%, and the loss of the remaining attribute data is small. Due to the integrity of data mining conclusions, these false data need to be filled in the pre-processing stage.

TABLE II. DISTRIBUTION OF MISSING DATA BY ATTRIBUTE

Missing Data Attribute	The Number of/a	Percentage of Missing Data /%
MMDIRL (P)	574	1.67
Load (Q_{ch})	216	0.66
The outlet temperature of chilled water (T_{eo})	168	0.56
The inlet temperature of chilled water (T_{ei})	173	0.45
Cooling water outlet temperature (T_{co})	305	0.87
Cooling water inlet temperature (T_{ci})	71	0.34

B. Abnormal Data

Abnormal performance and data transmission cannot be avoided when devices such as multimedia multi-dimensional information villages and sensors are in extreme action. In data preprocessing, it is necessary to identify and clean these abnormal data.

C. Multidimensional Data

There are many design variables related to multi-dimensional information on rural landscapes. According to relevant theories and empirical models, the influencing factors of multimedia in multidimensional information rural manufacturing are attributed to five factors: load (Q_{ch}), the

outlet temperature of cold water (T_{eo}), immersion temperature of cold water (T_{ei}), the outlet temperature of cold water (T_{ci}) and immersion temperature of cold water (T_{co}). There are several specific influencing factors. Whether the dimension of data continues to decline needs to be analyzed in the pre-processing stage.

At the same time, in the process of data sorting, there are fewer problems such as redundancy of points, repeatability of fields, lack of data accuracy, and format contradiction so that it is screened and processed in a simple, fast, and one-sided.

Given the above data quality analysis, this project must conduct data preprocessing, focusing on abnormal data and multidimensional data. It is the following primary research work.

V. MISSING DATA FILLING BASED ON THE DT LEARNING

DT learning algorithm is a classification algorithm based on analogy learning, which learns by comparing the given check object data with similar training data [14-15]. The missing-value processing step of the DT learning algorithm compares a sample data set with labels to determine the correlation between the data and classification. The features of the sample set are compared to the newly entered unlabeled data, and the classification labels that correspond to the most similar features are extracted. This approach ensures that the data is correctly classified. Generally speaking, only the k most similar data in the first part of the data set are selected, and the most common classification among the k most similar data is selected as the new data classification to realize the filling of missing values.

A simple data analysis of Table I shows that the original running data has a continuous data missing phenomenon. As shown in Table III, see lines 34145-34149 for missing data. The missing attribute is the multi-dimensional information rural landscape (P) value, while the other attributes are complete. The kNN method populates the data to get the missing attribute values.

The classic process of taking advantage of kNN: First, have exclusive properties. ($P, Q_{ch}, T_{ci}, T_{co}, T_{ei}, T_{eo}$) For each missing data, select $k=3$, and finally close to ". Distance quote: The known data values are populated as unknown attributes. Get all unknown data values. Through kNN program processing, the values of the unknown data P in the data in lines 34145-34149 are shown below.

TABLE III. MISSING DATA IN THE ACTUAL RUNNING DATA

S/N	P/kW	Q_{ch} /kW	T_{ci} /°C	T_{co} /°C	T_{ei} /°C	T_{eo} /°C
34145	-	515.1	24.3	28	17.1	12.4
34146	-	483.0	29.8	33.7	13.4	9
34147	-	515.1	23.1	28	16.8	12.1
34148	-	496.7	24.5	33.7	13.3	8.8
34149	-	494.7	27.6	33.7	13.6	9.1

The DT learning algorithm is set for similar processing for other missing values in the original data.

VI. ABNORMAL DATA IDENTIFICATION AND CLEANING BASED ON THE DT LEARNING ALGORITHM

DT learning algorithm analysis, the essence of the principle is to collect similar things, not similar things, into distinct kinds of processes. Data set 'D' contains 'n' objects, generated clusters 'K', and the clustering algorithm divides them into clusters EQ. (7) and EQ. (8).

$$k(k \leq n) C_1, \dots, C_k \quad 1 \leq i, j \leq k \quad C_i \subset D \quad (7)$$

$$C_i \cap C_j = \emptyset \quad (8)$$

The DT learning algorithm identifies data outliers by assuming that objects belong to large, dense, small, sparse, or none of these clusters. An overview of the above ideas as a general method of outgroup point (outlier) identification is as follows.

As illustrated in Fig. 2, a data object is regarded as an outlier if it does not fit into any of the clusters that have been created.



Fig. 2. a is abnormal data.

Outliers are depicted in Fig. 3 and occur when a data object is located far from the center of the cluster to which it belongs.



Fig. 3. a and b are abnormal data.



Fig. 4. a and b are abnormal data.

As seen in Fig. 4, if a data object is included in a minor or sparse cluster, then every object included in that cluster is considered an outlier.

In general, if the abnormal data does not belong to any cluster, as shown in Fig. 2, or belongs to a smaller cluster than the others, as shown in Fig. 4, the cluster data is deleted and ignored (even if the deletion does not affect the data sample capacity). As shown in Fig. 3, abnormal data must be cleaned correctly if it is far from the cluster center. In this paper, based on the DT learning algorithm, the cluster subset is clustered, and the abnormal data is replaced with the central value of the subset group to which the abnormal data belongs.

In order to study the relationship between multi-dimensional information about rural landscapes and a single variable, it is necessary to classify the data in detail. When the rural load rate of multimedia multidimensional information is fixed (66%) and the immersion temperature of cooling water (28.3°C), the rural landscape of multimedia multidimensional information constitutes a changing 2D array. The total number of data groups is 176, as shown in Table IV.

The DT learning algorithm was applied to perform cluster analysis on the above data, and the number of categories $k=4$ of the DT learning algorithm was selected to achieve the clustering results, as shown in Fig. 5.

Based on the above abnormal data identification principle, the cluster results in Fig. 5 are analyzed, and it is found that there are abnormal data in Fig. 5, which requires sub-cluster data cleaning and processing.

Take the cluster in the upper left corner of Fig. 5. The data set comprises 56 groups of data objects and data objects. (7.6,58.4), (7.5,58.4) are abnormal data (separated from the center of each cluster), and the tree learning algorithm of the cluster object is re-determined to take the number of cluster categories $k=4$. If cluster results, data object (7.5,58.3), (7.6,58.5) and other owning cluster center is (8.2,58.3), and cluster center (8.1,59.1) replaces (7.5,58.5), (7.6,58.4) and other abnormal data, abnormal data is cleaned.

Similarly, the data in the lower left, upper right, and right corners of Fig. 5 is cleaned using a similar determination tree learning algorithm. Fig. 6 shows the results of re-clustering the cleaned data. Comparing Fig. 6 vs. Fig. 5, it is found that the data distribution is centralized, which is a good clustering result.

TABLE IV. SHOWS THE ACTUAL RUNNING DATA OF THE COUNTRYSIDE WITH MULTI-DIMENSIONAL MULTIMEDIA INFORMATION

S/N	P/kW	Q_{ch} /kW	T_{ci} /°C	T_{co} /°C
1	59.5	354.6	29.5	8.7
2	60.6	354.7	29.6	8.8
5	59.7	354.7	29.6	8.9
175	60.8	354.7	29.6	11.2
176	61.1	354.7	29.6	11.2

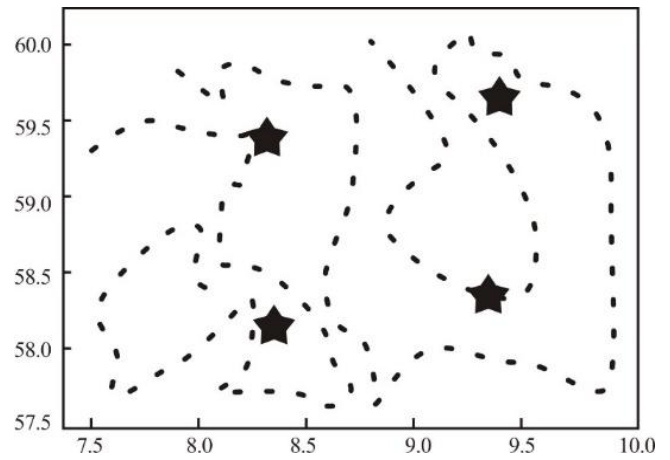


Fig. 5. Results of DT learning algorithm for raw data.

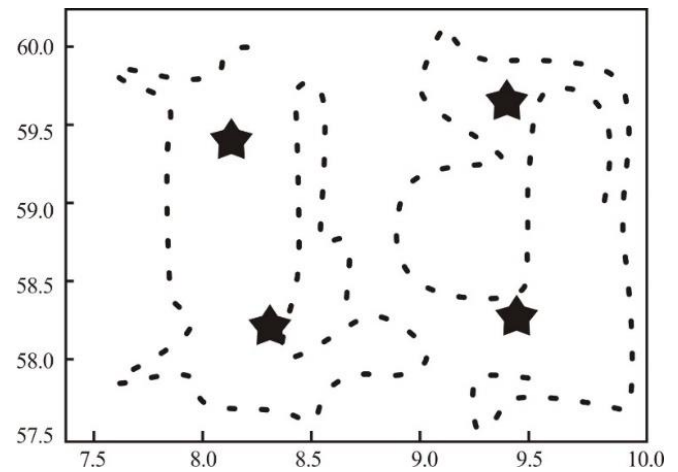


Fig. 6. Re-clustering results after cleaning the abnormal data in Fig. 5.

VII. DATA DIMENSION REDUCTION BASED ON THE DT LEARNING ALGORITHM

Principal Component Analysis (PCA) is a method of multiple statistical analysis that uses a linear transformation to reduce the number of different indexes to a small number of comprehensive indexes uncorrelated. By analyzing the characteristics of the matrix, the original data is projected in linear space to reduce the dimension of the data. The central principle is based on each component's cumulative dispersion contributions to determine the principal component.

The core process of principal component analysis is as follows: (a) The covariance matrix of the eigen-centralization matrix. (b) Computing the eigenvector associated with the covariance matrix's eigenvalue is necessary. (c) Calculate the dispersion contribution rate of each component. (d) Calculate the cumulative dispersion contribution rate.

In the core process, primary element analysis accumulates dimensional scattered data contribution rate to a manually set limit. This data influence represents the entire data set and creates high-order relegation data.

The most original problem of regression research is to study the relationship between multi-dimensional information

rural landscape (P) and load (Q_{ch}), the outlet temperature of cold water (T_{eo}), immersion temperature of cold water (T_{ei}), the outlet temperature of cold water (T_{co}), immersion temperature of cold water (T_{ci}) and other factors. From the perspective of variables, it is related to six-dimensional variables. If we dig down all six variables, there may be some insignificant factors. It will do useless work and will waste many computing costs. This section used the PCA method to analyze six-dimensional variables to obtain the main influencing factors. The results of the DT learning algorithm are shown in Table IV.

The results of the DT learning algorithm in Table V are analyzed, showing that the cumulative dispersion contribution rate of the upper four variables reaches 99.93%. Since the dispersion contribution rates of the first four components and the last two components show a significant difference in the number of digits, the first four components are the main components, and the six-dimensional data in this data set is maintained at four-dimensional data. Through the follow-up data, freezing waters found that there is a direct relationship between the multi-dimensional information rural landscape (P) and load (Q_{ch}), the outlet temperature of cold water (T_{eo}), and the immersion temperature of cold water (T_{ci}), and the corresponding mathematical relationship is obtained. The rationality and applicability of data degradation of the DT learning algorithm have been proved here.

TABLE V. RATE OF VARIANCE CONTRIBUTION AND CUMULATIVE RATE OF VARIANCE CONTRIBUTION

S. No. of Composition	Rate of Variance Contribution (%)	Rate of Cumulative Variance Contribution (%)
1	83.951	82.851
2	17.758	99.309
3	2.014	100.023
4	1.47	100.193
5	1.304	100.197
6	1.303	100

VIII. CONCLUSION

Given the characteristics of multi-dimensional information data in rural landscape design, the general steps and methods of data preprocessing applied in some other fields cannot be directly applied to the big data preprocessing of buildings. The data preprocessing system of multi-dimensional information rural landscape design suitable for multimedia and the corresponding data quality problem processing methods are presented.

DATA AVAILABILITY

On request, the corresponding author will provide access to the data used to support the findings of this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING STATEMENT

This research study is sponsored by these projects: Project one: The Young Innovative Talents Project (Humanities and Social Sciences) of Guangdong Province, the project number is 2019WQNCX127. Project two: Philosophy and Social Sciences "the 13th Five-Year Plan" Youth Project of Guangdong Province, the project number is GD20YYS05. Project three: 2020-2021 academic year scientific research project of Guangdong Peizheng College; the project number is pzxjzd02.Thank these projects for supporting this article!

REFERENCES

- [1] K. Rajala, M. G. Sorice & V. A. Thomas, "The meaning(s) of place: identifying the structure of sense of place across asocial-ecological landscape", *People and Nature*, vol. 32, no. 9, pp. 310-316, 2020.
- [2] A. A. Chibilev & V. P. Petrishchev et al., "The soil-ecological index as an integral indicator for the optimization of the land-use structure", *Geography and Natural Resources*, vol. 37, no. 4, pp. 348-354, 2016.
- [3] Q. Zhang, S. Wang, & K. Jian, "Design of campus wetland landscapes: a case study of the bailuxi wetland of University of Sanya in Sanya, Hainan, China", *Journal of Landscape Research*, vol. 55, no. 11, pp. 393-399.
- [4] Z. Wang, "Study on the multimedia application in college aerobics teaching: a learning interactive perspective", *Revista De La Facultad De Ingenieria*, vol. 32, no. 2, pp. 759-767, 2017.
- [5] J. Kim, "Subdivision design and landscape structure: case study of the woodlands, texas, us", *Urban Forestry & Urban Greening*, vol. 38, pp. 232-241.
- [6] Y. Chen, "Coastal mountain landscape and urban plant planning based on remote sensing imaging", *Arabian Journal of Geosciences*, vol. 14, no. 8, pp. 1-15, 2021.
- [7] Costanza & Robert, "Whole watershed health and restoration: applying the Patuxent and GWYNN Falls landscape models to designing a sustainable balance between humans and the rest of nature", *Journal of Landscape Research*, vol. 8, no. 2, pp. 333-344, 2015.
- [8] Y. U. Ziping, "Landscape design of the Ming Dynasty outer city wall in Nanjing from the perspective of experience tourism", *Journal of Landscape Research*, vol. 10, no. 6, pp. 137-139, 2018.
- [9] Corlett & T. Richard, "The role of rewilding in landscape design for conservation", *Current Landscape Ecology Reports*, vol. 1, no. 2, pp. 127-133, 2016.
- [10] R. Oppermann, E. Aguirre, R. Bleil, J. D. Calabuig & A. Schraml, "A rapid method for monitoring landscape structure and ecological value in European farmlands: the Lisa approach", *Landscape Online*, vol. 90, pp. 1-24, 2021.
- [11] M. Lin, T. Lin, L. Jones, X. Liu & X. Lu, "Quantitatively assessing ecological stress of urbanization on natural ecosystems by using a landscape-adjacency index", *Remote Sensing*, vol. 13, no. 7, pp. 727-733, 2021.
- [12] L. Z. Bian, "Analysis principles of landscape design small city", *Applied Mechanics and Materials*, pp. 744-746, 2015.
- [13] M. B. Lott, "Collective memory as an aesthetic landscape: the costume and scenic design of William Shakespeare's as you like it", *Dissertations & Theses – Gradworks*, vol. 8, no. 1, pp. 11-15, 2015.
- [14] L. Valetti, A. Pellegrino & C. Aghemo, "Cultural landscape: Towards the design of a nocturnal lightscape", *Journal of Cultural Heritage*, vol. 12, no. 21, pp. 200-203, 2019.
- [15] Woudstra & Jan, "Designing the garden of GEDDES: the master gardener and the profession of landscape architecture", *Landscape & Urban Planning*, vol. 2, no. 1, pp. 1-7, 2018.

Intelligent Detection System for Electrical Equipment based on Deep Learning and Infrared Image Processing Technology

Mingxu Lu¹, Yuan Xie²

College of Artificial Intelligence Applications, Shanghai Urban Construction Vocational College, Shanghai, 201415, China¹
Library, Shanghai Construction Engineering School, Shanghai, 200241, China²

Abstract—The demand for the reliability of power grid systems is gradually increasing with the development of the power industry. And it is necessary to promptly identify and eliminate the hidden dangers. To meet the needs of online monitoring and the early warning of electrical equipment, an intelligent detection system based on deep learning and infrared image processing technology is proposed in this study. Firstly, the infrared image is preprocessed for noise reduction. Then, an improved SSD (Single Shot MultiBox Detector) network is used to optimize the infrared image detection method. Based on this, an intelligent detection system for electrical equipment is designed. The results show that the mAP value of the improved SSD network after 1200 iterations is about 92.58%, and its area under the Precision Recall (PR) curve is higher than other algorithms. The simulation analysis results of the detection system show that the improved method detects a fault degree of 57.85%, which is closer to the 59.74% in the real situation. The experimental results indicate that the newly established intelligent detection system for electrical equipment can effectively detect its abnormal situations.

Keywords—Deep learning; infrared images; electrical equipment; intelligent detection; adaptive median filtering

I. INTRODUCTION

With the rapid development of the power industry, the requirements for the reliability of the power grid system are becoming increasingly high. Equipment fault detection is a key to intelligent detection for electrical equipment. Electrical equipment faults have randomness, periodicity, concealment, and multiple occurrences, requiring constant attention to the status of electrical equipment [1-2]. Online monitoring and security warning have become important functional requirements for the power grid system. However, in order to timely identify problems and eliminate hidden dangers, a large amount of manpower and objects need to be consumed, so the intelligence and automation of the power grid are gradually being put on the agenda. The infrared image detection method is a very effective online monitoring method. It can not only detect defects through online detection, but also be combined with other methods. It can locate the fault and bring great convenience to maintenance [3]. In recent years, deep learning technology has made rapid progress. And more and more image recognition tasks have achieved good performance in deep learning solutions. It is increasingly applied to various industries, greatly promoting industry reform and innovation. Image processing, as an important branch, is becoming

increasingly intelligent with the promotion of neural network. Simultaneously, deep learning can improve the accuracy and efficiency of neural network image feature extraction and classification [4]. The long-term development of the power grid system has accumulated a large amount of infrared detection data, which can be applied to artificial intelligence technology. Based on image classification technology and automatic analysis and processing of electrical images, the current difficulties in manual data management can be effectively overcome, reducing on-site measurement and post maintenance workload. Therefore, achieving intelligent recognition of temperature anomalies in infrared images is a necessity for the development of power systems. Improving the efficiency of monitoring and ensuring the smooth operation of the power grid also inevitably requires the large-scale application of such automation technologies. Research has shown that deep learning techniques based on neural networks can improve the accuracy of image processing [5]. The combination of methods can improve the accuracy of equipment fault detection. However, improving the accuracy of equipment fault detection can easily lead to a decrease in the efficiency of the method. In order to meet the needs of online monitoring and early warning of electrical equipment, an intelligent detection system based on deep learning and infrared image processing technology was proposed in this study. Firstly, preprocessing such as noise reduction was performed on the infrared image. Then, an improved SSD (Single Shot MultiBox Detector) network was used to optimize the infrared image detection method. On this basis, an intelligent detection system for electrical equipment was designed. It is hoped to further improve the practical application effect of electrical equipment testing methods.

II. RELATED WORKS

The maintenance of power equipment is related to the safety of production, so it is necessary to improve the accuracy and effectiveness of equipment fault detection methods [6]. The abnormal situation detection of device can be judged using information from infrared images. In power equipment fault detection, infrared images can reflect the basic information of the equipment. Wang et al. designed an online electrical equipment fault detection method based on infrared images and developed relevant software. This method can perform state evaluation based on real-time device status, and its effectiveness has been proven in practical applications [7]. After obtaining the infrared image of the device, DL

model can classify the obtained image features for fault diagnosis and analysis [8]. In Shen et al.'s study, they used DL for feature extraction and classification after feature extraction of infrared images. The improved method can improve the accuracy of the detection method [9]. In the fault detection of power equipment, researchers have designed a fault detection method based on infrared images. And this method uses DL and infrared images for defect recognition and classification. Finally, through the combination of methods, the reduction of operation and maintenance costs was achieved, and the efficiency of fault detection was improved [10]. Siah et al. showed that DL combined with this technology can play a role in detecting abnormal situations in the examination of respiratory equipment. Based on infrared imaging technology, they can use infrared images to analyze whether there is an air leak in the respirator. This can effectively control the widespread spread of the virus and reduce the likelihood of public health crises [11].

The infrared images can effectively reflect the basic situation of the device. But during the image acquisition, the clarity of the image may be reduced due to the noise. To reduce the noise impact on image clarity, researchers used filtering algorithms for fault detection, which showed high image processing performance. The filtering algorithm can perform segmentation preprocessing on the collected images, thereby improving the quality of the detected image. In the study by Jayanthi et al., the improved filtering algorithm can improve the quality of collected images for tumor detection. The results confirm that the new method can assist in tumor detection and improve the accuracy of patient diagnosis [12]. The combination of median filtering and clustering algorithm can realize the image super pixel segmentation. This method can reduce noise impact and improve the quality of infrared image acquisition. Infrared images can effectively reflect the abnormal situation of device after segmentation [13]. In relevant research, filtering algorithms can significantly reduce noise impact after being weighted. And this method combines mean algorithm, which effectively protects the details of the image, improves the information processing ability and the detection performance of the image [14]. To obtain more fault information, Zhang et al. proposed the introduction of DL technology to improve filtering algorithm. This new method can mine more time-related data information. In fault recognition simulation experiments, this method can effectively extract fault features and perform accurate classification [15]. In DL, Single Shot MultiBox Detector (SSD) is a high-precision algorithm that can be used for image information processing. After continuous method improvements, Leng et al. were able to fuse features from different levels during the image sampling process. At the same time, the uniform generation of image anchors improves the accuracy of detection while ensuring the efficiency [16]. In the study by Zhou et al., SSD can use residual networks for image feature extraction when detecting sample targets. This method exhibits stronger target detection ability and higher detection accuracy in performance comparison [17].

In the above research, infrared images, filtering algorithms, and SSD have shown good performance in object detection, image processing, and feature extraction. And the combination

of methods improves the equipment fault detection accuracy. However, improving equipment fault detection accuracy can easily lead to a decrease in the efficiency of the method. To improve this situation, after preprocessing the infrared image, the improved SSD was used for image processing. It is hoped to further improve the practical application effect of electrical equipment testing methods.

III. INTELLIGENT DETECTION SYSTEM FOR ELECTRICAL EQUIPMENT BASED ON DL AND INFRARED IMAGE PROCESSING TECHNOLOGY

A. Construction of an Intelligent Detection System for Electrical Equipment and Research on Infrared Image Preprocessing

To ensure the long-term stable operation of electrical equipment detection system, system design should follow the principles of safety, adaptability, practicality, maintainability, and scalability. The framework of detection system is Django, which mainly includes the model, template, and view layer. Business logic is one important core in this framework, including infrastructure, application, model layer, etc. According to the actual needs of electrical equipment testing, this system is divided into four management modules: infrared image recognition, inspection tasks, image management, and personnel management in Fig. 1.

Infrared image recognition is the core part of the detection system, which requires preprocessing of infrared images to achieve functions such as temperature extraction, device recognition, and anomaly detection. Periodic inspections are required in inspection tasks management, as well as re-inspection and timely warning of any abnormal equipment. In image data management module, personnel can transmit image samples and view the running status and related data. The permission management of staff is the main function of the personnel management module, which can view the basic data information of relevant personnel.

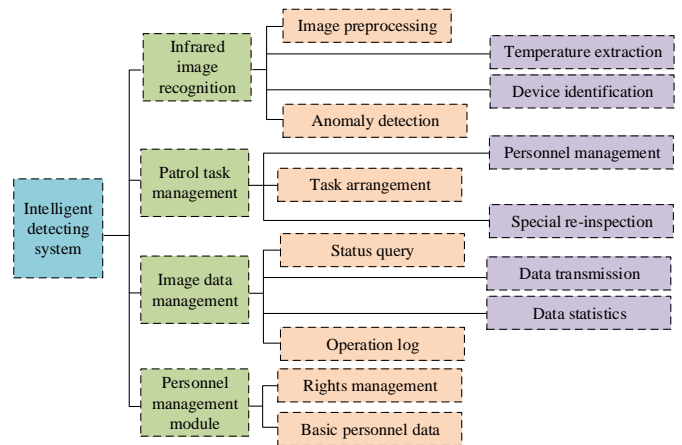


Fig. 1. Functional architecture of electrical equipment intelligent detection system.

In Fig. 2, the infrared image recognition module mainly includes image reading, image preprocessing, temperature extraction of electrical equipment, infrared image segmentation, and electrical equipment recognition. In this

study, the infrared image recognition module focuses on the noise processing of electrical equipment infrared images, as well as extraction and classification of image features. In response to the issue of infrared images being susceptible to factors such as noise, a series of processing measures were carried out. It has improved the research content of the infrared image recognition module in the intelligent detection system of electrical equipment.

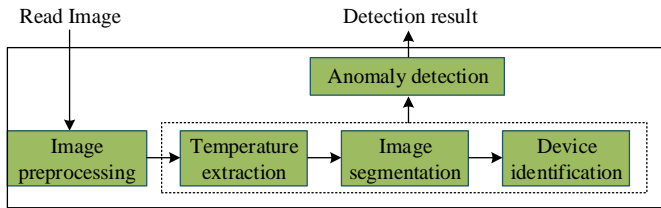


Fig. 2. Infrared image recognition module.

Noise characteristic analysis is one of the main characteristics of infrared image characteristics, and image noise has a certain degree of randomness and regularity. Noise has a significant impact on the detection of electrical equipment images, mainly including background noise, detector noise and amplifier noise, etc. The infrared image can be denoised by using Gaussian filter, mean filter, median filter, and other methods. The median filter can calculate the results in linear time. This method is simple in calculation and can perform fast filtering processing. The noise processing process of median filter algorithm includes the following contents. First, all pixel values in the neighborhood of pixel to be processed are arranged into a sequence according to the size of the pixel value. The pixel value at the interval position is the desired median. Then, the pixel values that need to be processed are replaced with median values to improve the closeness between the neighboring pixel values and the true values in Fig. 3.

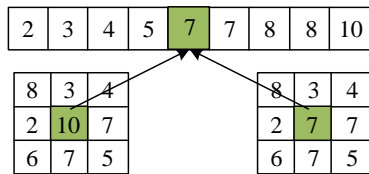


Fig. 3. Schematic diagram of the principle of median filtering.

In the median filtering algorithm, a window is defined that it is mostly odd with a length of $L = 2N + 1$, and N is a positive integer. The sample signal $\{X(i - N), \dots, X(i + N)\}$ can be obtained. Formula (1) is the output value of the median filter.

$$Y(i) = MED\{X(i - N), \dots, X(i), \dots, X(i + N)\}, i \in Z \quad (1)$$

In formula (1), $X(i)$ is the value of window center signal sample, $N = (L - 1) / 2$, and Z represents the set of integers. The median filtering template is larger, the noise removal effect is better, but the clarity of the infrared image decreases. To effectively remove noise and ensure the details in infrared images, researchers propose an AMF. This method can automatically select template size and select the smallest

template for filtering processing when corresponding pixel value noise removing. It can simultaneously remove noise and ensure image clarity. In AMF, the window corresponding to the image pixel (x, y) during filtering is first defined as S_{xy} , and then the following two processing processes are performed. The first step is the processing of the first layer in formula (2).

$$Z_{A1} = Z_{med} - Z_{min}, Z_{A2} = Z_{max} - Z_{med} \quad (2)$$

In formula (2), Z_{min} , Z_{med} , and Z_{max} represents the minimum, median, and maximum values of pixel grayscale in window S_{xy} , respectively. If $Z_{A1} > 0$ and $Z_{A2} < 0$, it needs to be transferred to the second algorithm layer. On the contrary, it is necessary to increase the size of window S_{xy} . If the size is less than or equal to the maximum window size allowed by S_{xy} , then the processing of the first layer needs to be repeated. Otherwise Z_{med} will be output. Formula (3) is the processing method for the second layer.

$$Z_{B1} = Z_{xy} - Z_{min}, Z_{B2} = Z_{max} - Z_{xy} \quad (3)$$

In formula (3), Z_{xy} is the grayscale value at pixel point (x, y) . If $Z_{B1} > 0$ and $Z_{B2} < 0$, the current Z_{xy} is output, otherwise Z_{med} is output. AMF can be used to filter out salt and pepper noise. During image processing, it needs to preserve image details as much as possible while removing noise. After denoising, infrared images become blurry, making it difficult for subsequent feature extraction and recognition. Therefore, image enhancement processing is also required after denoising [18]. The grayscale histogram of infrared images can provide information such as the overall contrast, average brightness, and dynamic range of image pixel values. To enhance the clarity of infrared image gray histogram, piecewise linear transformation, gamma correction and histogram equalization can be used to enhance the clarity of image details. In this experiment, the adaptive histogram equalization method is selected for image enhancement. The main improvements of this method in image processing include two points. Firstly, the grayscale threshold of the histogram is set, and the parts exceeding the threshold are cropped. And it is evenly divided into various gray levels to avoid excessive enhancement of noise points. Secondly, interpolation methods are used to accelerate the equalization of grayscale histograms. After the equalization of adaptive histogram, the gray histogram of infrared image can enhance image contrast without obvious noise enhancement.

B. Intelligent Detection Algorithm based on DL and Infrared Images

After preprocessing the infrared image, it is necessary to extract and classify the image features. SSD can classify the feature maps of the infrared image according to their size for target detection at various scales. When conducting intelligent detection of electrical equipment, the first step is to input the infrared image of the electrical equipment. After the feature extraction through convolutional layers, feature maps with different scales are generated. Then, prior boxes of different

scales are generated on these special maps with different scales, and the predicted target boundary boxes are detected and classified. On this basis, redundant detection boxes are suppressed and deleted, and finally detection results are

generated. The method of suppressing detection boxes is processed using non-maximum values. Fig. 4 shows the flowchart of the SSD algorithm.

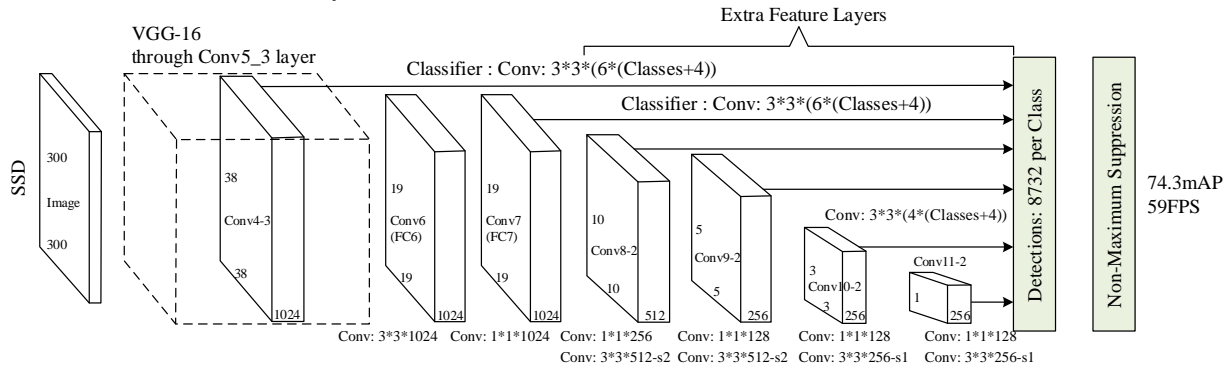


Fig. 4. Flowchart of SSD.

The length and width of prior boxes number and prior boxes in SSD need to be set in advance. These parameters can significantly affect the efficiency and accuracy of algorithm [19]. In response to the characteristics of electrical equipment fault detection, appropriate prior boxes need to be set, so SSD can be improved. The improved SSD first utilizes the K-means algorithm to iteratively analyze data targets and find the optimal number of prior boxes. Due to the fact that Euclidean distance is a commonly used method in K-means, this method can increase the computational error of the larger prior boxes. The intersection over union (IOU) of two prior boxes is selected as the standard for method judgment in formula (4).

$$IOU = \frac{Detection \quad Result \cap Ground \quad Truth}{Detection \quad Result \cup Ground \quad Truth} \quad (4)$$

In formula (5), *Detection Result* is the predicted bounding box and *Ground Truth* represents the true bounding box. IOU value reflects the proximity between the predicted bounding box and the true bounding box, which is proportional to each other. Formula (5) is the Euclidean distance calculation based on IOU as the standard.

$$d(box, centriod) = 1 - IOU(box, centriod) \quad (5)$$

In formula (5), *box* is the prior box and *truth* represents the true box. Formula (6) is the objective function of K-means.

$$S = \min \sum_{i=1}^k [1 - IOU(box - truth)] \quad (6)$$

When $k=4$, the clustering objective function remains basically stable, so the number of prior boxes is set to 4. After determining the setting of prior box parameters, improvements are made to the basic network. VGG-16 is used for extracting features, which has high testing accuracy but high computational complexity. To reduce the computational difficulty and improve the computational speed, researchers proposed the MoblieNet network structure [20]. MoblieNet

network can convert ordinary convolutions into a combination of deep convolutions and point convolutions using deep separable convolutions. Formula (7) is the calculation of deep separable convolutions.

$$F \times F \times D \times N \times N + 1 \times 1 \times D \times K \times N \times N \quad (7)$$

In formula (7), F means convolution kernel dimension, D is input depth, N refers to input width and height, and K represents output depth. By comparing deep separable convolutions computational complexity with that of standard convolutional networks, the comparison result representation method in formula (8) can be obtained.

$$\frac{F \times F \times D \times N \times N + 1 \times 1 \times D \times K \times N \times N}{F \times F \times N \times N \times D \times K} = \frac{1}{K} + \frac{1}{F^2} \quad (8)$$

If 3×3 convolutional kernel is used in the calculation, the computational difficulty can be reduced by 8-9 times. The MoblieNetV2 network structure uses 1×1 convolutional layer to expand the number of feature map channels. Then, the feature extraction of infrared images was carried out, and 3×3 deep separable convolution method is used. Next, the extracted image features were dimensionally reduced, and 1×1 convolutional layer was selected. In MoblieNet V2 network structure, the linear activation function is used at the low dimension layer, and the feature reuse structure of ResNet is introduced to improve the low dimension data collapse and the lack of reuse features in the MoblieNet V1 network. It can improve the accuracy of the algorithm and reduce latency. To make the network layer thinner, the width scaling factor α is introduced in the MoblieNet V2 network structure, which changes the depth of the input and output channels to αD and αK , respectively. In formula (9), the computational complexity of the MoblieNet V2 network can be reduced to α^2 of the original computational complexity.

$$F \times F \times \alpha D \times N \times N + 1 \times 1 \times \alpha D \times \alpha K \times N \times N \quad (9)$$

MoblieNet V2 network structure optimizes the network structure and activation function on the basis of MoblieNet V1. While improving testing accuracy, it can also save more feature information, reducing computational difficulty.

Therefore, in this experiment, the VGG-16 basic network of the SSD algorithm will be replaced with the MoblieNet V2 network structure. The improved SSD algorithm is developed to improve the algorithm's computational speed and testing accuracy.

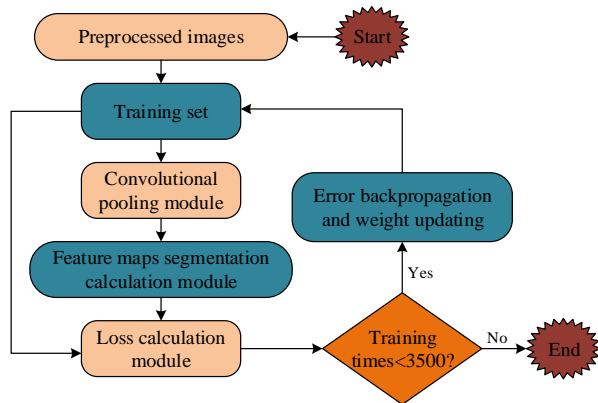


Fig. 5. Training flowchart of improved SSD algorithm.

Fig. 5 shows the training flowchart of the improved SSD algorithm. When algorithm training, the loss function uses the weighted sum of the position error loc and the confidence error $conf$ of SSD in formula (10).

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (10)$$

In formula (10), N means positive samples number, c represents the confidence value of the predicted category, l stand for the position of the predicted corresponding bounding box, g represents the position parameter of the true bounding box, and α refers to the weight coefficient. For loc , the $smooth_{L1}$ loss function in formula (11) is used.

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \quad (11)$$

From this, the calculation of L_{loc} in formula (12) can be obtained.

$$L_{loc}(x, l, g) = \sum_{i \in Pos} \sum_{m \in \{cx, cy, w, h\}} x_{ij}^k smooth_{L1}(l_i^m - \hat{g}_j^m) \quad (12)$$

In formula (12), cx, cy are the center coordinates of the positioning box, w, h are the width and height of the positioning box, and \hat{g} is the value obtained after encoding. Formula (13) is the encoding method for central coordinate.

$$\begin{cases} \hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx}) / d_i^w / variance[0] \\ \hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy}) / d_i^h / variance[1] \end{cases} \quad (13)$$

In formula (13), d represents the prior box position, and

$variance$ is hyperparameter, and \hat{g} can be scaled. Formula (14) is the encoding method for width and height.

$$\begin{cases} \hat{g}_j^w = \log(g_j^w / d_i^w) / variance[2] \\ \hat{g}_j^h = \log(g_j^h / d_i^h) / variance[3] \end{cases} \quad (14)$$

For $conf$, the softmax loss function in formula (15) is used.

$$L_{conf}(x, c) = - \sum_{i \in Pos} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in Neg} \log(\hat{c}_i^0) \quad (15)$$

In formula (15), $x_{ij}^p \in \{1, 0\}$, which represents the parameter indicator. $x_{ij}^p = 1$ means that the predicted boundary box i is in a state of coincidence with the actual boundary box j . At this point, the category is p , and the higher the probability prediction, the smaller the loss. Therefore, the probability passes and softmax is generated.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Performance Verification Experiment of Intelligent Detection Algorithm

AMF can effectively reduce the interference of salt and pepper noise, and the processed infrared image quality can be improved with less external influence, making the process more convenient and faster. The denoising performance of AMF is compared with that of Gaussian filter, mean filter, and median filter algorithm. Signal to Noise Ratio (SNR), Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) are selected as performance comparison indicators. SNR and PSNR are larger, as well as MSE is smaller, algorithm's denoising effect is better. The deep learning framework is Tensorflow-1.13.0rc2. During network training, the initial learning rate is set to 0.0001, the beam size is 16, the encoding scaling factor variable = [0.1 0.1 0.1 0.2 0.2], and the weight coefficient of the loss function is $\alpha=0.2$. The results are shown in Table I.

From Table I, the SNR and PSNR of AMF are greater than those of Gaussian filter, mean filter, and median filter algorithms under different noise concentrations. The MSE of AMF is less than that of Gaussian filter, mean filter, and median filter algorithms. The performance comparison results of different noise processing methods show that AMF is better than Gaussian filter, mean filter, and median filter algorithms in infrared image noise processing.

In Fig. 6, the filtered images of different methods were compared when the noise concentration was 0.02. From the figure, AMF has a higher clarity of the filtered image and better similarity to the original image. Compared to other image processing methods, AMF exhibits better image processing performance. This method can effectively handle the impact of noise and has better denoising effect.

TABLE I. RESULTS OF DIFFERENT NOISE TREATMENT METHODS

Noise concentration	Mean filtering algorithm			Median filter algorithm		
	SNR/dB	PSNR/dB	MSE	SNR/dB	PSNR/dB	MSE
0.01	50.05	52.68	0.004	52.68	58.95	0.003
0.02	43.52	51.63	0.006	49.21	57.89	0.004
0.03	41.18	49.29	0.007	45.26	55.37	0.005
0.04	36.51	44.74	0.008	43.26	51.77	0.008
0.05	31.14	39.32	0.014	42.25	51.03	0.013
0.06	22.49	30.58	0.036	37.52	46.15	0.034
0.07	15.94	24.02	0.071	36.41	44.98	0.067
Noise concentration	Gaussian filter algorithm			AMF		
	SNR/dB	PSNR/dB	MSE	SNR/dB	PSNR/dB	MSE
0.01	53.7336	56.559	0.003	55.31	61.90	0.002
0.02	46.7262	55.437	0.004	51.67	60.78	0.003
0.03	44.217	52.9176	0.006	47.52	58.14	0.004
0.04	39.1986	48.0318	0.007	45.42	54.36	0.005
0.05	33.4356	42.2178	0.008	44.36	53.58	0.006
0.06	24.1434	32.8338	0.011	39.40	48.46	0.009
0.07	17.1156	25.7856	0.023	38.23	47.23	0.011

Mean average precision (mAP) can be used to evaluate the performance of network models. In the performance research of the intelligent detection algorithms based on DL and infrared images, mAP was used to evaluate the performance. Fig. 7 shows the mAP curve in validation set. From Fig. 7, as the number of training sessions increases, the mAP value of the proposed model continues to increase. In 0-1200 iterations,

the mAP value changes the most significantly and the accuracy increases significantly, indicating that the model is still in the learning stage in the 0-1200 iterations. When iterations exceed 1200, the mAP value gradually stabilizes and the curve gradually converges. After the model was trained in the validation set, its average accuracy was tested in the test set, and the final mAP value measured was 92.58%.

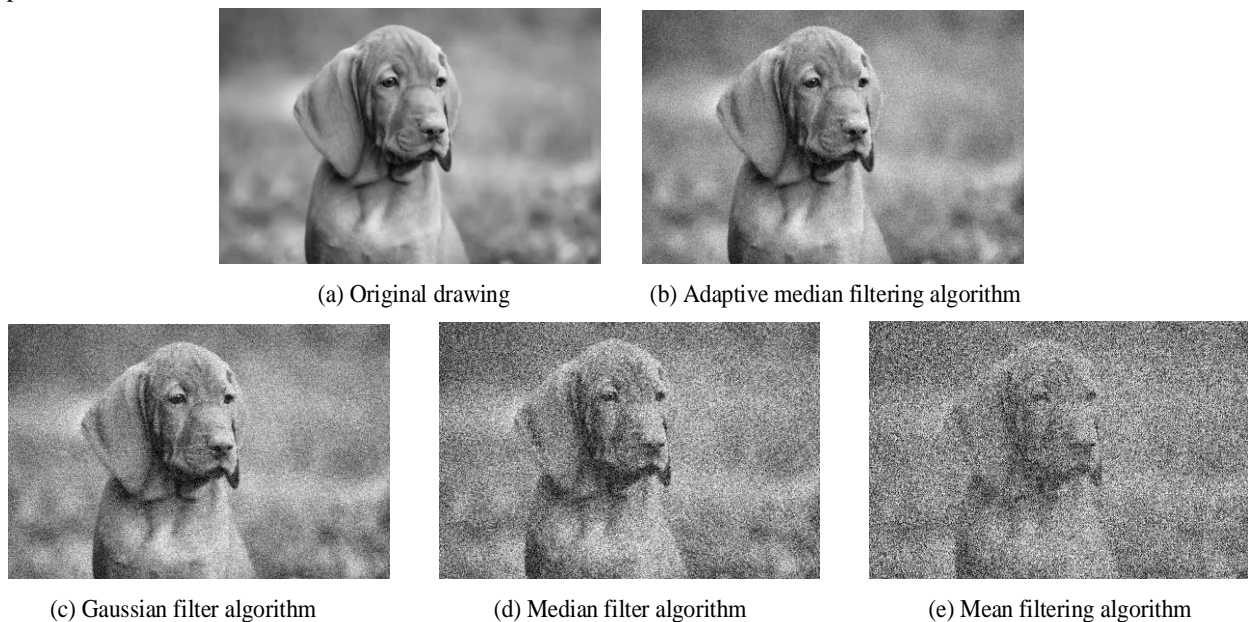


Fig. 6. Filter images of different algorithms.

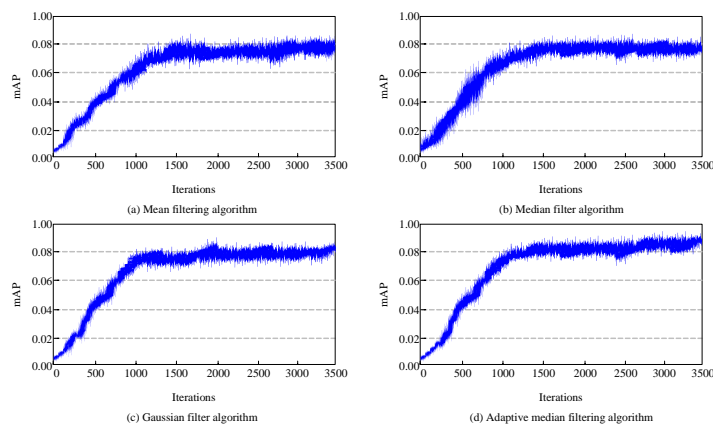


Fig. 7. mAP curve.

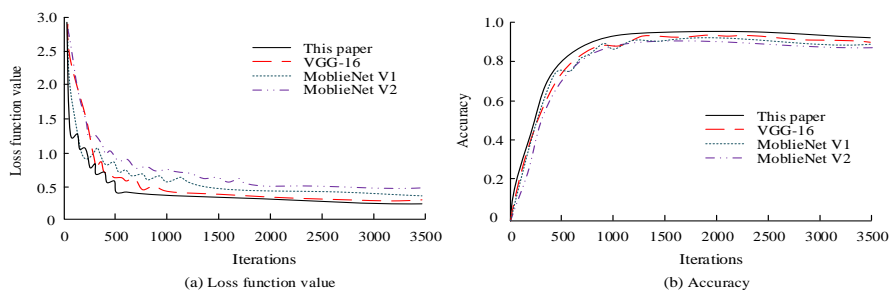


Fig. 8. Change of loss function value and quasi curvature.

The loss function can be used to describe the convergence degree of algorithm, and the accuracy can reflect algorithm precision. Therefore, this study will compare the performance of the proposed algorithm with VGG-16, MoblieNet V1, and MoblieNet V2 in Fig. 8. The algorithm in this paper tends to be stable after 500 iterations and reaches the convergence state. Its final loss function value is 0.31. Its accuracy stabilized after 950 iterations, and its final accuracy was 92%. Compared to VGG-16, MoblieNet V1, and MoblieNet V2, this algorithm has higher convergence speed and accuracy.

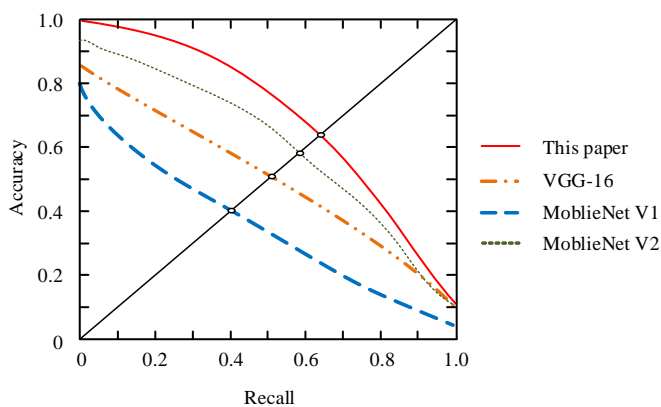


Fig. 9. Comparison of PR curves.

Precision (P) and Recall (R) are used to describe sample classification accuracy and have many applications in evaluating model performance. The Precision Recall (PR) curve is a comprehensive judgment of P and R. This study compared performance of our algorithm with VGG-16,

MoblieNet V1, and MoblieNet V2. The results are shown in Fig. 9. From Fig. 9, the area under PR curve of the proposed algorithm is higher than that of VGG-16, MoblieNet V1, and MoblieNet V2 algorithms, proving that the optimized SSD algorithm has good accuracy. The image feature extraction accuracy in improved algorithm is good, and the running time is shortened, which has good detection and classification performance.

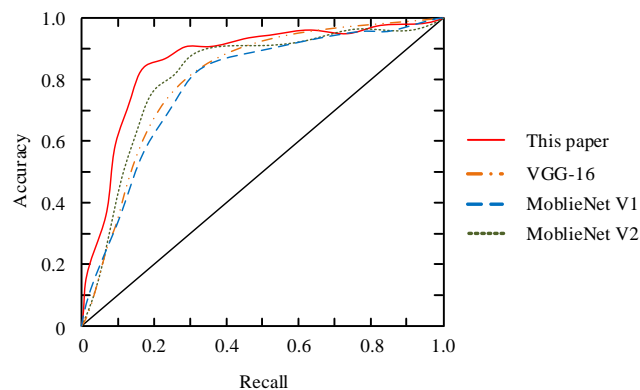


Fig. 10. Comparison of ROC curves.

In Fig. 10, the receiver operating characteristic curve (ROC) of different algorithms was simultaneously verified. From the figure, the area under ROC curve of the algorithm proposed in this experiment is the largest, and the validation results of ROC curve are consistent with those of PR curve. Based on the above performance analysis results, the algorithm proposed in this experiment has high accuracy, short running time, and good performance.

B. Simulation Analysis of Intelligent Detection System for Electrical Equipment

Circuit breakers in electrical equipment may be damaged if they are out of oil or in other abnormal conditions. In severe cases, explosions may occur and intelligent detection of circuit breakers is necessary. In the simulation analysis of the intelligent detection system for electrical equipment, the experiment selected the fault detection of circuit breakers for example analysis. Firstly, pre-processing such as noise reduction and image enhancement was performed on the collected infrared of the circuit breaker. Then, an improved algorithm was used for graphic feature extraction and classification. By comparing abnormal and normal images, the abnormal condition of the circuit breaker was analyzed and judged. To verify the superiority of the intelligent detection method for electrical equipment based on deep learning and infrared processing technology, the threshold method, region method, clustering method, and the methods in this experiment were compared for anomaly detection of circuit breakers. The results are shown in Fig. 11.

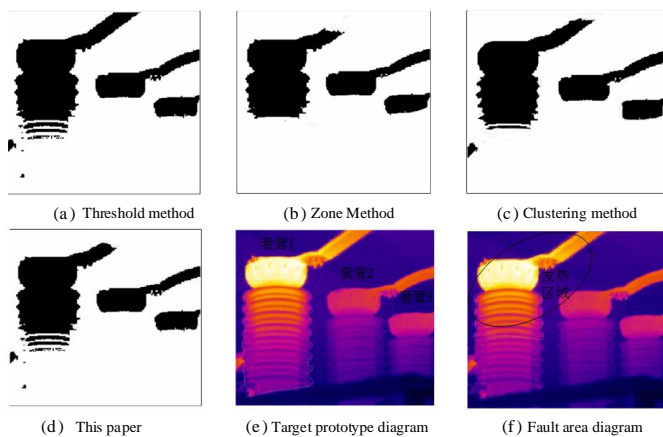


Fig. 11. Comparison of infrared image analysis of circuit breakers.

From Fig. 11, in the original image of the circuit breaker, there is a large heating area in sleeve 1. After the judgment and analysis, the contact of sleeve 1 has a heating situation, and there is also a problem of poor contact, resulting in abnormal heating of the sleeve. After comparing four detection methods, it can be seen that the improved infrared image is closer to the real image. At the same time, the fault degree of various detection methods was compared. The improved method detected a fault degree of 57.85%, which is closer to the 59.74% in the real situation. Its fault detection accuracy is higher. The intelligent detection system established in this experiment can be better applied to the fault detection of electrical equipment. In order to verify the effectiveness of the proposed method in detecting fault states, comparative experiments were conducted on a test set for 20 types of faults. At the same time, the fault detection rates of different detection methods [21-22] were compared, as shown in Table II. The best detection results for each type of fault in the table are highlighted in bold. The experimental results show that the proposed method achieved the best results 14 times, which is higher than other methods. At the same time, the proposed method has a detection rate of more than 10% higher than other methods for faults 16 and 19. The above

simulation results confirm that the proposed method exhibits high detection performance in electrical equipment fault detection.

TABLE II. RESULTS OF FAULT DETECTION RATES

Fault type	Reference 21	Reference 22	Reference 23	This paper
Fault 1	99.38	99.40	94.53	100.00
Fault 2	99.25	97.61	98.51	98.75
Fault 3	28.16	67.46	92.54	56.52
Fault 4	99.50	99.50	91.54	100.00
Fault 5	34.46	99.40	92.54	91.79
Fault 6	100.00	100.00	100.00	100.00
Fault 7	100.00	100.00	100.00	100.00
Fault 8	97.51	95.32	86.57	94.41
Fault 9	34.33	64.87	31.84	70.89
Fault 10	59.20	82.88	90.55	97.01
Fault 11	76.50	86.37	82.59	90.92
Fault 12	98.63	95.22	90.55	98.14
Fault 13	84.95	91.44	89.55	97.01
Fault 14	99.50	99.50	96.52	100.00
Fault 15	34.06	65.17	32.84	52.34
Fault 16	54.97	87.26	87.56	98.26
Fault 17	94.77	92.44	91.54	94.72
Fault 18	90.05	91.24	88.56	100.00
Fault 19	41.92	75.02	45.37	87.94
Fault 20	62.07	86.37	81.59	89.68
Average value	74.46	88.82	83.26	90.92
Optimal number	6	3	3	14

V. CONCLUSION

The electrical equipment detection system requires accurate status recognition and judgment of electrical equipment to ensure the safety of production practice. In this design of the electrical equipment intelligent system, deep learning technology and infrared image processing technology are used to identify and judge the abnormal situations of electrical equipment. The experimental results of noise processing shows that the SNR and PSNR values of AMF are higher than those of Gaussian filter, mean filter and median filter algorithms. The MSE values of AMF are lower than those of other algorithms. In the research of infrared image processing, when the iterations exceed 1200, the mAP value of the improved SSD gradually stabilizes and the curve gradually converges. After this model was trained in the validation set, the improved algorithm was tested for average accuracy in test set, and the final mAP value measured was 92.58%. The area under PR curve of improved SSD algorithm is higher than that of VGG-16, MoblieNet V1, and MoblieNet V2 algorithms, demonstrating its good accuracy. In actual fault detection experiment of electrical equipment, the fault degree of various detection methods was compared. The

improved method detected a fault degree of 57.85%, which is closer to the 59.74% in real situation, and its fault detection accuracy is higher. The intelligent detection system established in this experiment can be well applied to electrical equipment fault detection. This study is based on a deep learning system for anomaly recognition of electrical equipment, which can automatically recognize infrared images of equipment captured during inspection tasks. When the equipment is abnormal, a warning message is issued to remind the staff to inspect and repair it, so that problems can be detected in a timely manner. This can prevent larger grid failures and greatly reduce economic waste and human and material resources. At present, this method is still in secondary development and testing, and has not been widely applied in the power grid. The main reasons are as follows. Firstly, most of the samples of infrared devices come from daily infrared charged detection, which has problems such as a small sample size, inconsistent format, non-standard shooting, and lack of data labeling. These issues directly constrain the improvement of diagnostic accuracy of infrared devices based on big data. Moreover, the appearance of different devices varies, which can have an impact on the adaptability and accuracy of recognition and partitioning algorithms. At present, the deep neural network used for infrared image partitioning requires a large amount of computation, mainly relying on the computing power of the backend server. From a practical application perspective, there is an urgent need for real-time diagnosis in infrared detection sites. Real time diagnosis can provide more timely equipment defect prompts for professionals, avoiding unnecessary retesting and retesting. However, due to the large amount of image information and the large number of image files, the use of background diagnosis mode requires high transmission bandwidth. They increase the burden on communication links and network backend, and also reduce the reliability of the entire diagnostic system. Therefore, further improvements are needed in future research on this detection method.

REFERENCES

- [1] Kahlen J N, Andres M, Moser A. Improving Machine-Learning Diagnostics with Model-Based Data Augmentation Showcased for a Transformer Fault. *Energies*, 2021, 14(20): 6816-6835.
- [2] Wadi M. Fault detection in power grids based on improved supervised machine learning binary classification. *Journal of Electrical Engineering*, 2021, 72(5): 315-322.
- [3] K Wang, S Yuan, Z Yao, J Gao, J Feng. Design and Implementation of Infrared Image Classification Algorithm for Defective Power Equipment Based on Deep Learning. *Nonlinear Optics, Quantum Optics*, 2022, 56(1/2): 83-95.
- [4] J Wang, J Ou, Y Fan, L Cai, M Zhou. Online Monitoring of Electrical Equipment Condition Based on Infrared Image Temperature Data Visualization. *IEEJ Transactions on Electrical and Electronic Engineering*, 2022, 17(4): 583-591.
- [5] S Tiwari, K Falahkheirkhah, G Cheng, R Bhargava. Colon Cancer Grading Using Infrared Spectroscopic Imaging-Based Deep Learning. *Applied Spectroscopy*, 2022, 76(4): 475-484.
- [6] Bindhu V, Ranganathan G. Effective Automatic Fault Detection in Transmission Lines by Hybrid Model of Authorization and Distance Calculation through Impedance Variation. *Journal of Electronics and Informatics*, 2021, 3(1):36-48.
- [7] J Wang, J Ou, Y Fan, L Cai, M Zhou. Online Monitoring of Electrical Equipment Condition Based on Infrared Image Temperature Data Visualization. *IEEJ Transactions on Electrical and Electronic Engineering*, 2022, 17(4): 583-591.
- [8] Envelope W, Xr A. Electrical Fault Detection Equipment Based on Infrared Image Fusion. *Procedia Computer Science*, 2022, 208(1): 509-515.
- [9] Shen K, Shi Q, Wang H. Multimodal Visibility Deep Learning Model Based on Visible-Infrared Image Pair. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(6): 939-946.
- [10] K Wang, S Yuan, Z Yao, J Gao, J Feng. Design and Implementation of Infrared Image Classification Algorithm for Defective Power Equipment Based on Deep Learning. *Nonlinear Optics, Quantum Optics*, 2022, 56(1/2): 83-95.
- [11] C Siah, S Lau, S Tng, C Chua. Using infrared imaging and deep learning in fit-checking of respiratory protective devices among healthcare professionals. *Journal of nursing scholarship: an official publication of Sigma Theta Tau International Honor Society of Nursing*, 2021, 54(3): 345-354.
- [12] N Jayanthi, D Manohari, MY Sikkandar, MA Aboamer, MI Waly, C Bharatiraja. Multi-Model Detection of Lung Cancer Using Unsupervised Diffusion Classification Algorithm. *Computers, Materials and Continua (Tech Science Press)*, 2022, 31(2): 1317-1329.
- [13] Liu H, Hu J. An adaptive defect detection method for LNG storage tank insulation layer based on visual saliency. *Process Safety and Environmental Protection*, 2021, 156(1): 465-481.
- [14] Shao C, Kaur P, Kumar R. An Improved Adaptive Weighted Mean Filtering Approach for Metallographic Image Processing**. *Journal of Intelligent Systems*, 2021, 30(1): 470-478.
- [15] Zhang Y, Lv Y, Ge M. A Rolling Bearing Fault Classification Scheme Based on k-Optimized Adaptive Local Iterative Filtering and Improved Multiscale Permutation Entropy. *Entropy (Basel, Switzerland)*, 2021, 23(2): 191-213.
- [16] Leng J, Liu Y. Single-shot augmentation detector for object detection. *Neural Computing and Applications*, 2021, 33(8): 3583-3596.
- [17] F Zhou, F He, C Gui, Z Dong, M Xing. SAR target detection based on improved SSD with saliency map and residual network. *Remote Sensing*, 2022, 14(1): 180-189.
- [18] Zhang Q, Zhao L, Zhao L. A two-step robust adaptive filtering algorithm for GNSS kinematic precise point positioning. *Chinese Journal of Aeronautics*, 2021, 34(10): 210-219.
- [19] D Li, G Meng, Z Sun, L Xu. Autonomous multiple tramp materials detection in raw coal using single-shot feature fusion detector. *Applied Sciences*, 2021, 12(1): 107-113.
- [20] Wang Z, Feng J, Zhang Y. Pedestrian detection in infrared image based on depth transfer learning. *Multimedia Tools and Applications*, 2022, 81(27): 39655-39674.
- [21] Thomas J B, Shihabudheen K V. Neural architecture search algorithm to optimize deep transformer model for fault detection in electrical power distribution systems. *Engineering Applications of Artificial Intelligence*, 2023, 120: 105890.
- [22] Levin V M, Yahya A A. An innovative method of fault detection in power transformers. *International Journal of Electrical and Computer Engineering (IJECE)*, 2022, 12(2): 1123-1130.
- [23] Kullu O, Cinar E. A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults[J]. *Machines*, 2022, 10(11): 1105-1121.